Evolution of Western Palaearctic oak gallwasp communities

Richard Challis

Thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh

2007

Declaration

This thesis and the data presented in it are the result of my own research except where collaborative work has been duly acknowledged. The text does not exceed 100,000 words. No part of this work has been submitted to any other institution in application for a higher degree.

Abstract

This thesis has three major aims: (i) to utilise phylogenetic approaches to address a specific set of phylogeographic questions; (ii) to develop bioinformatic methods; and (iii) to improve understanding of the evolutionary history of the Western Palaearctic oak gallwasps and oak inquilines.

Phylogeographic studies typically address patterns of post-glacial range expansion from glacial refugia following the last ice age. However, since species' origins often predate the last ice age, longitudinal phylogeography attempts to reconstruct the historic patterns of colonisation that underlie refugial diversity. Understanding of events on this timescale can be used to guide conservation priorities by locating cradles of speciation and identifying the processes that affect genetic differentiation into populations, sub-species and sibling species. A review of the literature on Western Palaearctic phylogeography reveals that relatively few studies address the aims of longitudinal phylogeography. Within these studies, an emerging pattern of eastern origin of widespread Western Palaearctic taxa is identified and further investigated using the oak gallwasps as a model system. Eastern origins are identified in three widespread species of oak gallwasp, with a common timescale of origin approximately corresponding to the onset of the Pleistocene.

In order to address the aims of this thesis, protocols are established using existing methods and further bioinformatic methods are developed. Since longitudinal phylogeography is in its infancy, a protocol is established to provide a framework for future studies. Model-based trait mapping techniques are adopted for phylogeographic reconstruction, and a model reduction technique is developed that allows directions of longitudinal range expansion to be inferred. Given the potential importance of longitudinal phylogeographic concordance, a direct comparative method is proposed to allow quantitative comparison of intraspecific phylogenies. A further method is developed to allow consistent sets of molecular taxa to be

identified across multiple genes, allowing DNA barcoding to be applied to identify taxa in situations where data are missing for some genes in some samples, which should facilitate longitudinal phylogeography where morphological taxonomy is unresolved.

Western Palaearctic oak gallwasp communities have been extensively studied. This thesis resolves some of the outstanding issues in oak gallwasp and oak inquiline research. Cryptic lineages are identified in areas to the east of Europe, highlighting the importance of these areas as cradles of oak gallwasp diversity. The potential for human activity to alter longitudinal phylogeographic patterns is demonstrated for the oak gallwasp *Andricus kollari*, whose galls were historically important in trade. The scale of this trade is illustrated by the transfer of an entire phylogeographic clade into the UK from its region of origin to the east of the Mediterranean. Molecular taxonomy of the oak inquiline Synergini is shown to be inconsistent with the current morphological taxonomy, which will require extensive revision.

Acknowledgements

This thesis has been enhanced immeasurably by the many contributions, large and small – to my academic development, my general well-being, and my ideas about phylogenetics, phylogeography and gallwasps – from colleagues, collaborators, friends and family. To everyone who has helped along the way: thank you, this thesis could not have been completed without your generosity.

Firstly, thanks to Graham Stone, my supervisor and personal guide through the world of gallwasps. Thanks for having the patience to fill the occasionally persistent gaps in my knowledge of gallwasps, for the logistical support that has made this research possible and for all the suggestions and encouragement that have helped me to develop all those ideas into a thesis.

Thanks to all the members of the gall group, past and present for your enthusiasm in sharing ideas, gallwasps and data. In particular to Sonja Preuss for all the hours in the lab helping me to have enough data to answer many of the questions in this thesis. Thanks to Alex Hayward and James Nichols for the discussions, the samples and the sequence data; to Alex Aebi, Rachel Atkinson and Antonis Rokas for the samples and the lab work; and to Tracey Begg and Gordon Brown for the gallwasps.

Thanks also to the wider gallwasp community, in the UK and abroad, and in particular to all of the collaborators who made this thesis possible. Thanks to the Hungarians: to György Csóka for the samples, to Zsolt Pénzes for the comments on versions of Chapter 8, and especially to Zoltan Ács for generously sharing data and ideas on the evolution of the Synergini, and to George Melika for describing two of my cryptic lineages as new species – fulfilling my childhood ambition of discovering a new species. Thanks to Antonio Hernandez-Lopez for the samples and data and to Gil McVean, Serap Mutun, Jose-Luis Nieves-Aldrey, Rida Nuwayid, Olivier

Plantard, Juli Pujade-Villar, Alex Rowe, Ebrahim Sadeghi, Karsten Schönrogge and Majid Tavakoli for collecting gallwasps.

A special mention is also due to all the people that helped to collect samples that have not made it to the pages of this thesis. Thanks to David Challis for travelling with me around the eastern USA, subsisting primarily on Doritos; to John and Kathy Schick for sharing their home in California; and to the US National Park Service for help and advice.

Thanks to Mark Blaxter for an introduction to the world of MOTUs and for discussing the finer points of molecular phylogenetics, and to Richard Ennos and Andrew Rambaut for comments and advice on specific chapters.

Finally, thanks to all my family and friends for being there when I needed them and for gradually learning to appreciate the futility of asking when my PhD would be finished. Thanks Mum and Dad for trying to understand what I have been studying and, perhaps most incomprehensible of all, why. Most of all, thanks to Amy for all the love, support, advice and friendship along the way. Thanks for driving me around Spain and France to collect gallwasps, for patiently reading my work and helping me to get the ideas out of my head and onto paper.

Contents

Chapter 1

Introduction to the oaks, oak gallwasps and oak inquilines of the Western Palaearctic

1.1 Overview

The first aim of this thesis is to utilise phylogenetic approaches to address a specific set of phylogeographic questions: (i) in which direction and over what timescale have species' ranges expanded longitudinally from their geographic origins? (ii) in what ways and to what extent has human activity affected these patterns? and (iii) are patterns shared across species? The second aim is to develop bioinformatic methods to: (i) infer phylogeographic origins and directions of range expansion; (ii) assess equivalent barcodes across genes; and (iii) determine the impact of missing data on phylogenetic analyses of relatively small datasets. Both of these aims will be addressed using Western Palaearctic oak gallwasp communities as a model system. The third aim of this thesis is to improve understanding of the evolutionary history of the Western Palaearctic oak gallwasps and oak inquilines.

1.2 The Western Palaearctic

The unique geography and geological history of the Western Palaearctic makes it an excellent region in which to investigate longitudinal phylogeography. The geography, geological history and current phylogeographic understanding of the Western Palaearctic are outlined below.

1.2.1 Defining the Western Palaearctic

The Palaearctic zoogeographical region comprises Eurasia and North Africa and extends south as far as the Sahara and Himalayas (Figure 1.1). Although North Atlantic (55-34 million years ago, mya) and more recently Bering (9,000 years ago) land bridges have provided dispersal routes between the Palaearctic and Nearctic, the flora and fauna of the Palaearctic have typically been isolated from other regions by ocean, desert and mountain barriers. The Palaearctic region has been subdivided into

the Eastern and Western Palaearctic (Figure 1.1) along the Johannsen line, which follows the Yenisey River (de Lattin, 1967).

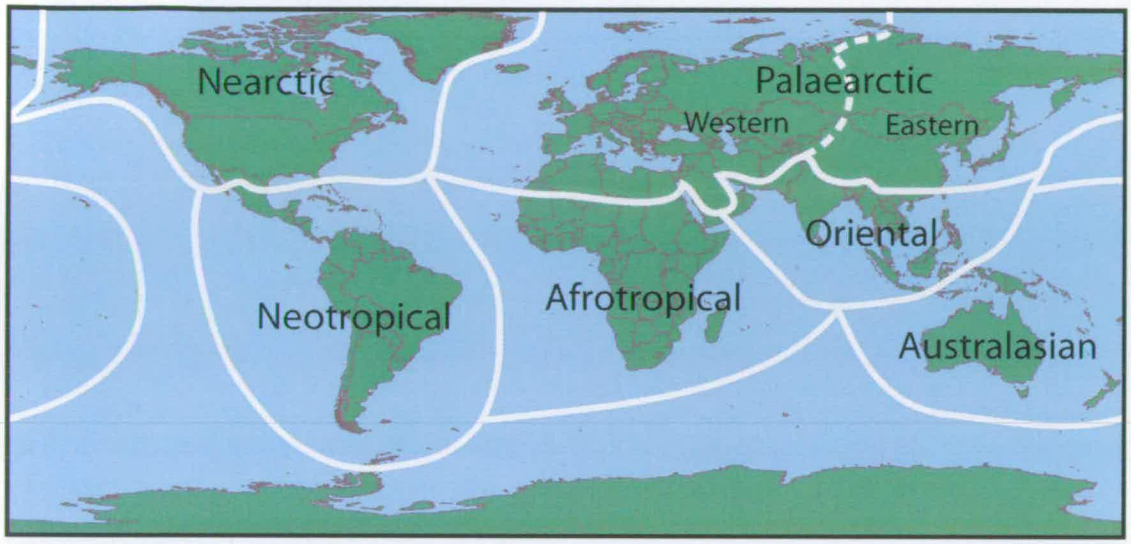


Figure 1.1 Zoogeographical regions of the world. The division between the Eastern and Western Palaearctic is also indicated.

1.2.2 Geological history

The recent geological history of the Western Palaearctic has been dominated by the glacial cycles of the Pleistocene (Table 1.1). During each ice age, large areas of the Western Palaearctic were covered by glaciers, with tundra and steppe habitat to the south only habitable to cold tolerant, boreal species. Most temperate species were restricted to the small areas of relatively mild climate that persisted in the extreme south of the region: in the Mediterranean peninsulas (Iberia, Italy, Balkans/Greece), south of the Caucasus Mountains, and in the Middle East. These areas were glacial refugia and provided sources of colonists to exploit newly available habitat to the north during the relatively warm, interglacial periods. The timing of interglacial periods were subject to regional variation

The number and duration of identified glacial periods varies between regions (Gibbard *et al.* 2005, *modified* 2007), complicating attempts to assign phylogeographic events (Section 1.2.3) to specific glacial or interglacial periods.

Table 1.1 Dates of recent geological stages and approximate numbers of glacial periods (Gibbard *et al.* 2005, *modified* 2007).

Series	Stage	Date (mya)	Number of glaciations
Holocene		0 – 0.012	0
Pleistocene	Late	0.012 – 0.12	1
Pleistocene	Middle	0.12 – 0.78	6
Pleistocene	Early	0.78 – 2.58	10

1.2.3 Phylogeography

The aim of phylogeography is to reconstruct historic processes within species by using information on the geographic distribution of genealogical lineages. Samples from across the range of a species can be interpreted in light of past geological processes (Section 1.2.2) to infer: colonisation routes; genetic bottlenecks; population expansion; division into populations and sub-species; and cryptic speciation.

The high density of research institutions and the relative ease of travel within Europe have lead to a historic sampling bias in Western Palaearctic phylogeography. Though the Western Palaearctic extends eastwards to the Yenisey River (de Lattin 1967), the vast majority of studies involve sampling in Europe with sampling strategies that appear to follow political rather than biogeographic boundaries. Many 'European' taxa, however, have distributions that extend eastwards beyond Europe, across the Western Palaearctic. Examples include climax woodland trees (sessile oak *Quercus petraea,* pedunculate oak *Q. robur*, common beech *Fagus sylvatica* and common alder *Alnus glutinosa*; Govaerts & Frodin 1998), mammals (brown bear *Ursus arctos* and house mouse *Mus musculus*; Mitchell-Jones *et al.* 1999), and birds (tawny owl *Strix aluco* and common blackbird *Turdus merula*; Hagemeijer & Blair 1997). Some of these widespread taxa have trans-Palaearctic distributions that extend into the Eastern Palaearctic (e.g. brown bear; Mitchell-Jones *et al.* 1999).

The majority of phylogeographic studies in the Western Palaearctic concern latitudinal changes in distribution associated with the contributions of glacial refugia to Holocene colonisation of Europe (Taberlet *et al.* 1998; Hewitt 1999, 2000, 2004). These studies focus on the identification of colonisation routes (e.g. Liukkonen-Antilla *et al.* 2002; Petit *et al.* 2002) and the identification of extant centres of genetic diversity, whether 'cradles' (centres of diversification) or 'museums' (reservoirs of diversity generated elsewhere) (Chown & Gaston 2000). Identification of such centres is increasingly viewed as an important component in assessing conservation priorities (Green 2005).

The Western Palaearctic is unique in having a longitudinal series of discrete refugia, which are isolated from each other by major mountain ranges oriented from east to west that create barriers to dispersal both into and out of the southern refugia. Many species show genetic differentiation between refugial populations, consistent with prolonged reproductive isolation, and refuge-specific polymorphism has been used to reconstruct recent postglacial range expansion routes in many taxa (Hewitt 1996, 1999, Taberlet *et al.* 1998). However, the barriers that have promoted genetic differentiation have apparently restricted post-glacial colonisation (the focus of latitudinal phylogeography), so while dispersal routes from refugia may be shared between species (Hewitt 1996, 1999), widespread patterns in the relative contributions of each refuge to northern populations have not been found (Taberlet *et al.* 1998), even among closely related species (Michaux *et al.* 2005).

In latitudinal phylogeographic studies, individual refugial populations are treated as sources of diversity from which post-glacial colonists are drawn. However, for widespread species that occupied multiple potential cradles and museums at the beginning of the Holocene, recent latitudinal range changes must necessarily have followed more ancient Pleistocene or pre-Pleistocene longitudinal dispersal events. The same barriers to dispersal that have prevented latitudinal phylogeographic concordance may be expected to preserve the genetic distinctness of individual refugia over multiple glacial cycles and may potentially have preserved

the original signature of longitudinal range expansion from the species origin. Reconstructing these more ancient longitudinal processes within a phylogeographic framework is important for two general reasons. Firstly, longitudinal phylogeography is central to understanding the relationships between populations and patterns in genetic diversity across the full distribution of widespread taxa (Ludt *et al.* 2004; Challis *et al.* 2007; Stone *et al.* 2007c). Secondly, comparison of patterns across species has the potential to allow reconstruction of relationships between similar communities in linked sets of cradles and museums, and hence to contribute to studies of community evolution. The field of biogeography is also concerned with the identification of species origins and dispersal routes (Brown & Lomolino 2004). However, since biogeographic studies are typically undertaken at the interspecific level, with limited intraspecific sampling, information about associated population processes is lost. The scope of longitudinal phylogeography therefore lies between existing approaches at the inter- and intraspecific levels.

Adoption of a longitudinal perspective makes it essential to consider the full longitudinal distribution of widespread species. The Western Palaearctic contains multiple major regional centres of endemism outside Europe. The Irano-Turanian region (which covers large areas of Turkey, Syria, Israel, northern Iraq, Iran and extends east into Asia; Takhtajan 1986) is a major speciation cradle, while the Caucasus (from northern Turkey, north between the Black and Caspian Seas) and the Ponto-Caspian regions (comprising the basins of the Black, Azov and Caspian seas) contain glacial refugia that are important museums of diversity (Davis 1965-1985, 1971; Bănărescu 1991; Chown & Gaston 2000; Orth *et al.* 2002). Within Europe, the Carpathian Basin and Mediterranean peninsulas have been identified as glacial refugia (Hewitt 1996, 1999, 2004) and areas of endemism (Médail & Quézel 1999) for a range of taxa. The significance of these regions as centres of biodiversity makes it likely that they also constitute major centres of intraspecific genetic diversity for widespread taxa. Excluding such regions from phylogeographic sampling may lead to erroneous conclusions on post-glacial colonisation routes, relationships between refugia, and possible source populations for invading or

introduced species. Human populations in Europe originated in the Fertile Crescent (the Levant, Syria and Iraq) and Turkey (Sokal *et al.* 1991), and sampling of populations from these regions is particularly important for taxa whose distributions are likely to have been altered by human activity – either directly through trade, or indirectly via accidental or commensal introduction. Important domesticated species in the Western Palaearctic, such as goats, have centres of origin in Asia and the Fertile Crescent but have little phylogeographic structure due to continued trade (Luikart *et al.* 2001). The house mouse *Mus musculus domesticus* did not follow the initial spread of agriculture across Europe but colonised from the Fertile Crescent once increased human impact on the environment created commensal niches during the first millennium BC (Cucchi *et al.* 2005).

The term longitudinal phylogeography reflects the distinction between latitudinal gradients of post-glacial colonisation of the Western Palaearctic and the underlying longitudinal processes of dispersal from each species' origin to colonise the refugial regions. While the principle that phylogeographic methods can be used to identify species origins and dispersal routes should be applicable beyond the Western Palaearctic, there are likely to be limitations to its use beyond the Northern Hemisphere. Phylogeographic studies in tropical regions, for example, typically consider both latitudinal and longitudinal dimensions (e.g. Aleixo 2006) while addressing questions that are regarded here as the subject of 'latitudinal' phylogeography.

1.3 Oaks

Oak trees are widespread throughout the Western Palaearctic and support diverse invertebrate communities, including oak gallwasps. The oak gallwasps (Section 1.4.1) are a group of parasitic Hymenopterans that are able to induce the formation of galls on oak trees and provide an excellent model system with which to address the aims of this thesis. Current understanding of oak taxonomy and distributions are outlined below.

1.3.1 Oak taxonomy and origins

The oak genus *Quercus* (Fagaceae; Fagoidea) is divided into two subgenera (Table 1.2): the Asian subgenus *Cyclobalanopsis* (cycle cup or ring cup oaks, 76 species), and the Holarctic subgenus *Quercus* (455 species) (Camus 1934-54; Nixon 1993; Govaerts & Frodin 1998). The subgenus *Quercus* is further subdivided into four sections based on DNA sequence data – *Quercus sensu stricto* (white oaks), *Cerris* (black oaks), *Lobatae* (red oaks) and *Protobalanus* (golden cup oaks) (Manos *et al.* 1999; Manos & Stanford 2001). Section *Cerris* has historically been divided into two species groups, *Ilex* and *Cerris* (Schwarz 1964), and this division has been supported by recent molecular studies (Manos *et al.* 1999; Bellarosa *et al.* 2005).

Table 1.2 Oak taxonomy and species richness.

Genus	Subgenus	Section	Group
Quercus	*Quercus*	*Cerris*	*Cerris*
Quercus	*Quercus*	*Cerris*	*Ilex*
Quercus	*Quercus*	*Lobatae*	
Quercus	*Quercus*	*Protobalanus*	
Quercus	*Quercus*	*Quercus sensu stricto*	
Quercus	*Cyclobalanopsis*		

The fossil record shows that evergreen oaks were present in Britain and continental Europe around 45 mya, during the middle Eocene (Brett 1960; Kvacek & Walther 1989); however, the current oak flora of the Western Palaearctic has more recent origins. Three of the present day oak sections, *Lobatae*, *Protobalanus* and *Quercus s.s.*, can each be recognized in Nearctic fossils from the Oligocene (20-35 mya) (Bones 1979; Manchester 1981; Muller 1981; Daghlian & Crepet 1983; Crepet & Nixon 1985). Palaearctic *Quercus s.s.* oaks are probably derived from ancestors that colonised from the Nearctic via the Bering Land Bridge during the Oligocene (Trelease 1924; White *et al.* 1997; Manos & Stanford 2001). The Nearctic and Palaearctic *Quercus s.s.* oak floras have had separate evolutionary histories for approximately 17 million years (Manos & Stanford 2001). Oaks in sections *Cerris* and *Quercus s.s.* reached the Western Palaearctic by spreading westwards along the

foothills of the Himalayas and the temperate highlands of Central Asia. The *Ilex* and *Cerris* groups in section *Cerris* probably spread westwards from Asia independently (Manos & Stanford 2001). Western Palaearctic oak radiations in both sections are relatively recent, and are thought to have been isolated from the Eastern Palaearctic oak fauna since the Pliocene (around 5 mya) (Manos & Stanford 2001).

1.3.2 Oak distributions

Section *Quercus s.s.* (white oaks) is found throughout the northern hemisphere, section *Cerris* (black oaks) is confined to the Palaearctic while the sections *Lobatae* (red oaks) and *Protobalanus* (golden cup oaks) are both confined to the Nearctic (Figure 1.2).

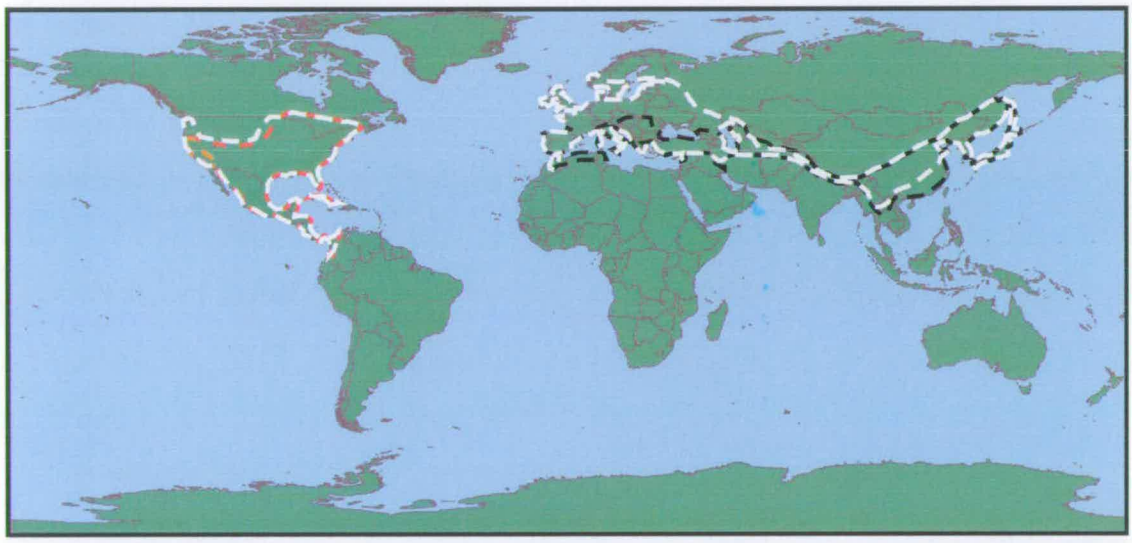


Figure 1.2 Distributions of oak sections *Quercus s.s.* (white), *Cerris* (black), *Lobatae* (red) and *Protobalanus* (gold).

1.3.3 Western Palaearctic oak flora

Camus (1934-54) recognised 76 oak species in the Western Palaearctic. Improvements in the understanding of genetic relationships between taxa have encouraged a more parsimonious approach, in which regional variants have been collapsed into a smaller number of widespread species. In a recent review of oak taxonomy, Govaerts & Frodin (1998) list only 29 Western Palaearctic oak species

(13 in section *Cerris*, 16 in section *Quercus s.s.*, see Table 1.3). The native distributions of the two Western Palaearctic oak sections can be described by two overlapping latitudinal bands (Figure 1.3, from Stone *et al.* 2007a).

Table 1.3 Western Palaearctic oak species, separated by Section and (within section *Cerris*) by group, following Govaerts and Frodin (1998).

Section	Group	Richness	Western Palaearctic species
Cerris	*Ilex*	5	*Q. alnifolia, Q. aucheri, Q. coccifera, Q. ilex, Q. rotundifolia,*
Cerris	*Cerris*	8	*Q. afares, Q. brantii, Q. castaneifolia, Q. cerris, Q. ithaburensis, Q. libani, Q. suber, Q. trojana.*
Quercus s.s.		16	*Q. canariensis, Q. congesta, Q. dalechampii, Q. faginea, Q. frainetto, Q. hartwissiana, Q. infectoria, Q. lusitanica, Q. macranthera, Q. petraea, Q. pontica, Q. pubescens, Q. pyrenaica, Q. robur, Q. sicula, Q. vulcanica*

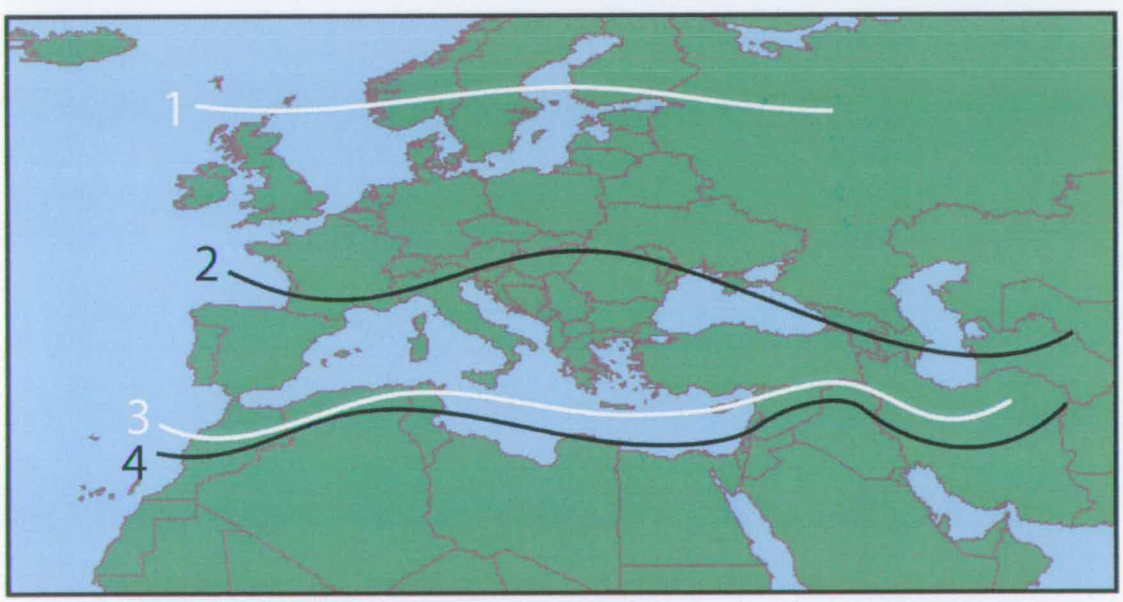


Figure 1.3 Regions of the Western Palaearctic occupied by the oak sections *Quercus sensu stricto* and *Cerris*. Line 1 represents the northern limit of oaks in the section *Quercus*, and of all oaks in the Western Palaearctic. Only section *Quercus s.s.* oaks are native between lines 1 and 2. Line 2 represents the northern limit of the natural distribution of section *Cerris* oaks. Oaks in sections *Quercus s.s.* and *Cerris* are naturally found together between lines 2 and 3. Line 3 represents the southern limit of oaks in the section *Quercus s.s.*. Only section *Cerris* oaks are found between lines 3 and 4. Line 4 represents the southern limit of section *Cerris* oaks. Distribution sources are as follows: Europe (Tutin *et al.* 1993), Turkey (Davis 1965-1985; Yaltirik 1982), the former USSR (Konarov 1936), Iran (Browicz & Menitsky 1971), Iraq (Townsend & Guest 1980), Palestine (Zohary 1966).

1.3.4 Glacial refugia

Divisions in the Western Palaearctic oak flora correspond with the ice age refugia, which form a longitudinal series of geographically separated regions (Stone *et al.* 2007b). Of the 29 Western Palaearctic species, only 5 are widespread. Twelve species are unique to Iran and the Caucasus, seven are unique to Iberia and five have very localised distributions (*Q. alnifolia*, for example is endemic to southern Cyprus). The pattern of distinction by refugial affiliation is illustrated by the species of the *Cerris* group of section *Cerris* which show a longitudinal series of partially overlapping distributions (see Figure 1.5) from *Q. suber* in northwest Africa and the Iberian Peninsula, through *Q. cerris* and *Q. trojana* in Italy, the Balkans, Greece and Turkey, to *Q. libani* and *Q. ithaburensis* in eastern Turkey, the Levant and western Iran, and *Q. castaneifolia* and *Q. brantii* along the Iranian Caspian coast and Zagros Mountains.

Estimates of the rate of northward range expansion by trees in Europe vary between 500 (Bennet 1986) and 2000 m per year (Huntley & Birks 1983), however, few Western Palaearctic oak species have naturally expanded their ranges beyond the glacial refugia since the last glacial maximum *c.* 18,000 years ago (Huntley & Birks 1983). *Q. pubescens* is now found throughout much of eastern and central Europe and throughout southern France to the Pyrenees but the most dramatic range expansion has been seen in *Q. petraea* and *Q. robur*. Chloroplast DNA haplotypes specific to the three European refugia have been used to investigate the origins of current northern populations of *Q. petraea* and *Q. robur* (Ferris *et al.* 1993, 1998; Dumolin-Lapegue *et al.* 1997; Petit *et al.* 2002; Figure 1.4). Both species are now found across Europe as far north as Finland, southern Scandinavia and Scotland in the west and the Central Russian Uplands in the east (Jalas & Suominen 1987; Tutin *et al.* 1993; Czerepanov 1995). Iberian populations are thought to have expanded into western France, and from there into the United Kingdom and north-eastwards into southern Norway. Italian populations expanded into central Europe as far north as southern Sweden, and a clear north-south contact zone between populations derived from the Iberian and Italian refugia now runs close to the border between

France and Germany (Ferris *et al.* 1993). Balkan populations expanded over a broad front, and lineages originating there are now found throughout central and western Europe, east into Ukraine and Byelorussia and north as far as southern Finland.

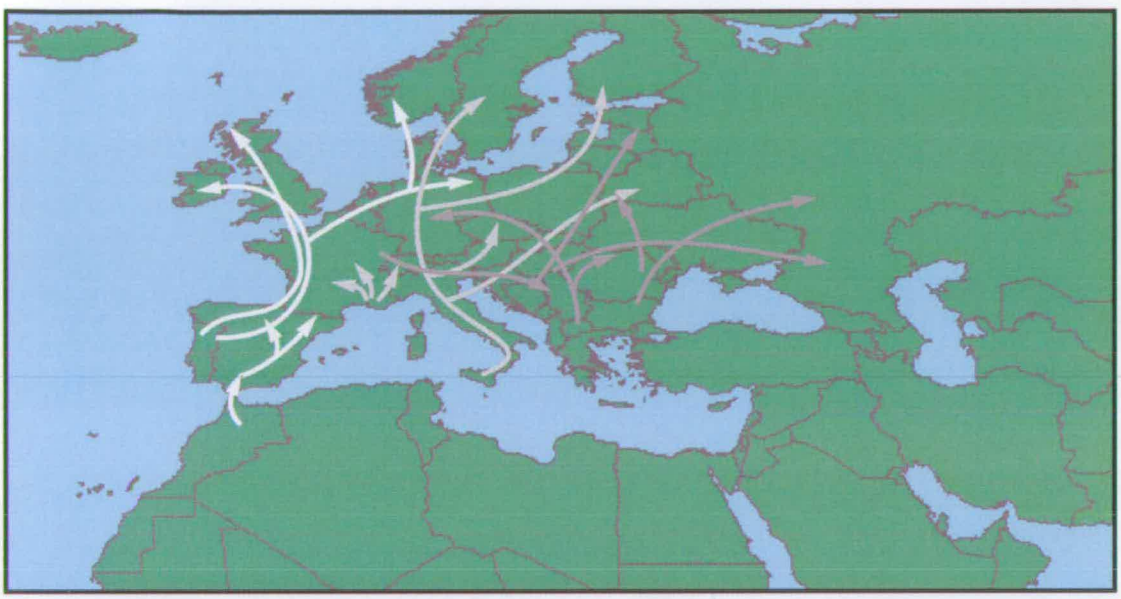


Figure 1.4 Contributions of alternative glacial refugia to northern oak populations in the current interglacial, after Petit *et al.* (2002). All of the arrows extending into northern Europe and eastwards into Russia are for *Q. petraea* and *Q. robur*.

1.3.5 Human-mediated dispersal

In addition to natural range expansion following glacial retreat, distributions of all the Western Palaearctic oaks have probably been affected by human activity. The most significant example of human dispersal of oaks in the Western Palaearctic is that of the Turkey oak, *Q. cerris*. This species is native to Europe south of the Alps and Carpathians, and the southwestern quarter of Turkey (Figure 1.5). Over the last 500 years, it has been planted extensively north and west of its native range, and has since become naturalised across much of northern and central Europe (Jalas & Suominen 1987; Tutin *et al.* 1993, Stone *et al.* 2001). The cork oak, *Q. suber*, though now predominantly an Iberian species, is also found in Italy (Figure 1.5). Italian populations may be the result of human introduction, however, it has been suggested that *Q. suber* may originally have spread westwards from an origin in the central Mediterranean or Middle East (Lumaret *et al.* 2005).

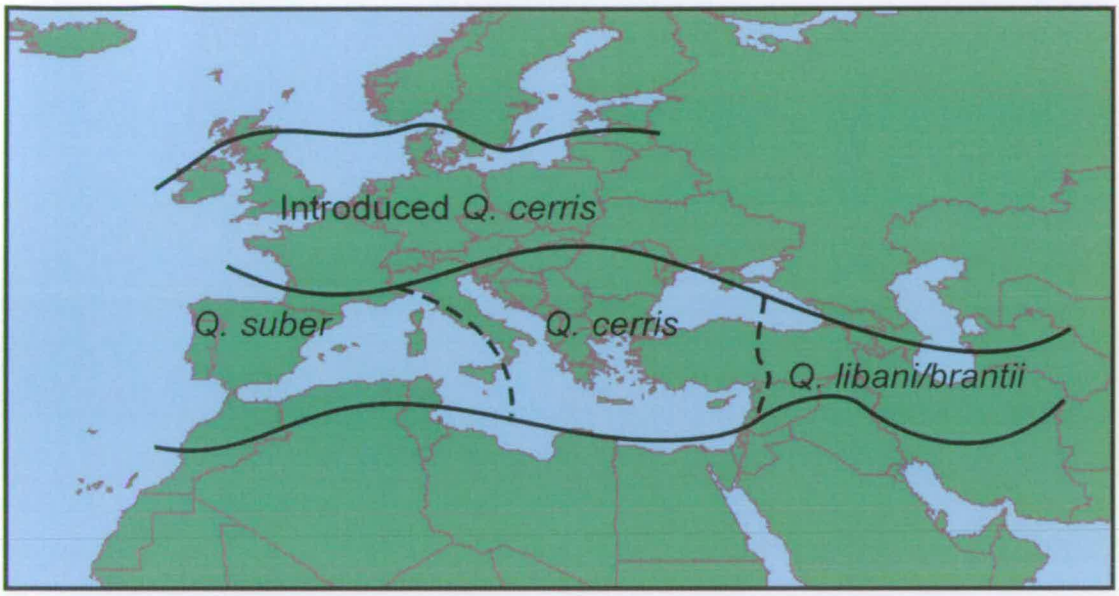


Figure 1.5 Approximate longitudinal divisions between the major Section *Cerris* oak species in the Western Palaearctic. Actual distributions are partially overlapping and *Q. suber* is rare in Italy. The eastern limits of the introduced range of *Q. cerris* in Russia, Belarus and Ukraine are unknown. After Stone *et al.* (2001).

1.4 Oak gallwasps

The oak gallwasps of the Western Palaearctic have been extensively studied and their unique biology makes them an excellent model system in which to investigate the potential for longitudinal phylogeographic concordance. Current understanding of relevant aspects of the biology of oak gallwasps is outlined below.

1.4.1 Introduction to the gallwasps

The gallwasps (family Cynipidae) are a monophyletic group of parasitic Hymenopterans within the superfamily Cynipoidea. There are around 1350 currently recognised gallwasp species (Table 1.4), although Nordlander (1984) has estimated that the actual number may lie between 3000 and 6000. The species richness of the Cynipidae is likely to rise as new species continue to be described in poorly-studied regions of high potential richness (Melika & Abrahamson 1997a, 1997b, 2000; Melika *et al.* 1999; Melika & Stone 2001). However, this increase is likely to be

offset as taxa currently recognised as separate species are shown to be the two generations (see Section 1.4.3) of a single species (Rokas *et al.* 2003b).

Table 1.4 Classification, diversity and host associations of Cynipinae, modified after Liljeblad & Ronquist (1998) and Ronquist (1999).

Tribe	Genera	Species	Host plant taxa
Aylacini	20	134	Asteraceae, Rosaceae, Lamiaceae, Papaveraceae, Apiaceae, Valerianaceae, Brassicaceae, one species on *Smilax* (Smilaceae)
Cynipini	25	*c.* 1000	Fagaceae (mostly *Quercus*, also *Castanea*, *Castanopsis* and *Lithocarpus*)
Diplolepidini	2	62	*Rosa* (Rosaceae)
Eschatocerini	1	3	*Acacia*, *Prosopis* (Fabaceae)
Pediaspidini	2	2	*Acer* (Aceraceae), *Nothofagus* (Fagaceae)
Synergini	8	151	Inquilines predominantly in galls induced by *Diastrophus* on *Rubus*, *Diplolepis* on roses and Cynipini on oaks. Also found in cecidosid moth galls on *Rhus* (Anacardiaceae)
Total	**58**	***ca.* 1350**	

All extant Cynipid gallwasps have been placed in the sub-family Cynipinae, which is further divided into six tribes which form two main trophic groups: the gall-inducers, and the gall-associated inquilines. Both groups are obligate plant parasites that develop within galls induced on specific plant tissues. The gall inducers redirect plant development to form structurally differentiated galls whose inner tissues are nutritive, and whose outer structures are protective (Stone *et al.* 2002b, Stone & Schönrogge 2003). Inquilines are obligate inhabitants of the galls of others; though able to induce the development of nutritive tissues in such galls, they cannot induce galls *de novo* (Stone *et al.* 2002b). The Aylacini (herb gallwasps, 134 species) mainly induce galls in herbaceous plants. Morphological data suggest that the Aylacini are not a monophyletic group, but represent a basal paraphyletic assemblage which gave rise to two major monophyletic lineages (Ronquist 1995, 1999). One lineage, the 'woody rosid gallers' (Ronquist 1999) consists of four tribes whose members all induce galls on woody rosaceous plants: the Diplolepidini (rose gallwasps), the Pediaspidini (sycamore gallwasps), the Eschatocerini (acacia

gallwasps) and the Cynipini (oak gallwasps). The oak gallwasps (tribe Cynipini) are the most species-rich group of gallwasps, with a global species richness of around 1000 species in 25 genera (Ronquist 1999, Pujade-Villar *et al.* 2001, Stone *et al.* 2002b). A second lineage gave rise to the tribe Synergini, whose members are all inquiline inhabitants of the galls of other gall-inducers, primarily Cynipids (Section 1.5).

1.4.2 The Western Palaearctic oak gallwasps

The Western Palearctic contains 139 recognised oak gallwasp species in 10 genera. The majority of Western Palaearctic oak gallwasps belong to four species-rich genera: *Andricus* (81 species), *Plagiotrochus* (14), *Neuroterus* (12), and *Cynips* (9). The remaining 23 recognised species are divided among 7 species-poor genera: *Aphelonyx*, *Biorhiza*, *Callirhytis*, *Dryocosmus*, *Pseudoneuroterus* and *Trigonaspis*. Asia Minor, the Caucasus and Iran represent major centres both of oak and oak cynipid gallwasp diversity (Rokas *et al.* 2003b, Atkinson *et al.*, 2007; Challis *et al.* 2007, Stone *et al.* 2007c), and new species continue to be discovered in these regions (Melika *et al.* 1999; Melika & Stone 2001, Melika *et al.* 2007).

1.4.3 Oak gallwasp life cycles

Oak gallwasp life cycles provide a rare example of cyclical parthenogenesis among the Metazoa (Folliot 1964; Askew 1984; Hebert 1987). Uniquely, cyclical parthenogenesis among the oak gallwasps involves strict alternation between the two reproduction modes, usually with the entire life cycle completed in one year (Askew 1984). Within a population there are two types of parthenogenetic generation females: (i) androphores produce haploid eggs by meiosis that give rise to male offspring, while (ii) gynephores produce diploid eggs that give rise to female offspring (Folliot 1964). Sexual generation females must mate to produce the next generation of parthenogenetic females, completing the life cycle (Figure 1.6). The wasps of the two generations often differ dramatically in size – the sexual generation adults are usually smaller and carry fewer eggs than their parthenogenetic generation

counterparts (Sanderson 1988), and develop far more rapidly within galls that are structurally different and usually smaller. Parthenogenetic generation wasps typically overwinter in galls induced in mid-late summer, while the smaller sexual generation wasps often induce gall development on ephemeral structures, such as catkins, in the spring (Askew 1984).

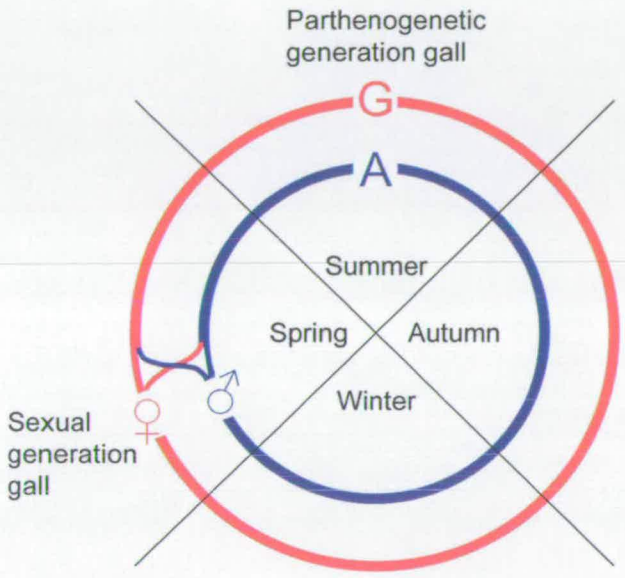


Figure 1.6 The life cycle of *Neuroterus quercusbaccarum* (after Askew 1984; Folliot 1964) highlighting the role of gynephores (G) and androphores (A) in maintaining cyclical parthenogenesis.

1.4.4 Gall morphology

The Cynipini induce the most morphologically diverse and complex galls of any insect group (Cornell 1983) and this diversity is achieved through modifications to the outer gall tissues (Stone & Schönrogge 2003). Certain gall morphologies are specific to particular host organs, for example, the artichoke galls of *Andricus fecundator* develop through modification of bud scales (Stone & Cook 1998). Many other gall morphologies develop on a range of organs, for example simple spherical galls can develop on leaves (*Cynips quercusfolii* pathenogenetic generation) or on catkins (*Neuroterus quercusbaccarum* sexual generation). Other gall shapes include conical galls, such as *Andicus conificus* (Stone & Cook 1998), and those with surface

spines, such as *Andricus lucidus* (Stone & Sunnucks 1992). Gall morphology can also be altered by changes in the texture of the inner and outer gall tissues. Species such as *Andricus kollari* have hardened gall tissues or thicker gall walls (Stone & Cook 1998), while at the opposite extreme, reduction in gall tissues can create air spaces within the gall (e.g. *A. quercustozae*). Modifications to the epidermis include surface hairs, bright coloration and secretions to attract ants or increase surface stickiness (Abe 1992; Stone *et al.* 2002b).

Three mutually compatible theories have been proposed to explain the adaptive benefit of gall morphologies: the Nutrition, Microenvironment and Enemy hypotheses (Stone & Schönrogge 2003). Living either partially or completely enclosed in plant tissues increases nutrient availability, protects gallwasps from environmental stress (Fay *et al.* 1993) and makes them less apparent to generalist predators. Although the gall protects oak gallwasp larvae from generalist predators, a number of species of specialist parasitoid wasps (103 in Western Palaearctic communities) are able to attack the larva developing within the gall and parasitoid species richness is high in oak gallwasp communities relative to non-galling relatives (Stone & Schönrogge 2003).

Of the three hypotheses, it is widely believed that only the Enemy hypothesis describes sufficient selective pressures to explain the diversity of structures that may be found on the same organ of the same host in the same location. Among gallwasps, ant tending (in which ants are attracted by a sugary secretion from the gall in return defending the gall from parasitoids; Washburn 1984; Abe 1992; Siebert 1993; Fernandes *et al.* 1999), and gall wall toughness (Abe 1997) have been proven to have adaptive benefits. Harder gall walls can only be penetrated by parasitoids with reinforced ovipositors. Many other apparently adaptive qualities have yet to be tested thoroughly, however, ways in which they could reduce parasitoid attack rates have been proposed. Fine hairs may interfere with parasitoid movement on the gall surface. Thick spines may hamper the efforts of larger predators, such as birds, to break into the gall. In addition to recruiting ants, secretions may be used to increase

the surface stickiness of galls so parasitoids are trapped on the gall for longer, increasing their vulnerability to predators. Larger galls, with increased distance between the outer gall surface and the inner wall of the larval chamber are only vulnerable to parasitoids with longer ovipositors or those attacking earlier in development when the nutritional reward is smaller. The presence of air spaces within the gall reduces the support along the length of a parasitoid ovipositor that would usually prevent it bending.

In some species, gall morphology varies across the species' range. The parthenogenetic generation galls of *Andricus quercustozae*, for example, have two distinct morphotypes: those found in Europe are brown and lack a sticky surface coating when mature (although those found on the Iberian peninsular are red and sticky during early stages of development), while those in Turkey and Iran (known as *Andricus insana*) are red and more spherical with a sticky surface coating. Detailed morphological analysis of the adult gallwasps that induce the two gall forms has revealed differences, which some regard as worthy of classification as a separate species, *A. insana* (Pujade-Villar *et al.* 2002), while others treat these regional differences as intraspecific variation (Melika *et al.* 2000; Melika & Bechtold 2001b). Genetic evidence to date has been insufficient to separate the two groups as discrete monophyletic lineages (Rokas *et al.* 2003a). Under the assumption that both gall morphologies are induced by members of a single species, there are at least two possible hypotheses to account for regional variation: (i) regional variation in gall morphology may reflect regional variation in selection pressure; and (ii) some gall morphologies may not be possible on alternate oak hosts, so regional variation in gall morphology may reflect regional variation in host oak species (Rokas *et al.* 2003a).

1.4.5 Host specificity and oak associations

The oak host associations of Western Palaearctic oak gallwasps fall into three main groups (Table 1.5): (i) both generations induce galls on oaks in section *Quercus s.s.*, (ii) both generations induce galls on oaks in section *Cerris*, and (iii) obligate alternation between parthenogenetic generations on one oak section and sexual

generations on the other (host alternation). The three exceptions (Stone *et al.* 2007a) are the parthenogenetic generations of *Andricus quercusradicis* and *Neuroterus anthracinus* which gall members of both oak sections and *Neuroterus albipes* which predominantly galls oaks in section *Quercus s.s.*, but the parthenogenetic generation galls are also recorded from section *Cerris* oaks.

Table 1.5 Life cycle types of the Western Palaearctic oak gallwasps showing the number of species per genus sharing the same sexual (♀♂) and parthenogenetic (♀) generation hosts.

Genus	♀♂ gen. host	♀ gen. host	Richness
Biorhiza	*Quercus s.s.*	*Quercus s.s.*	1
Cynips	*Quercus s.s.*	*Quercus s.s.*	9
Trigonaspis	*Quercus s.s.*	*Quercus s.s.*	5
Andricus	*Quercus s.s.*	*Quercus s.s.*	23
Neuroterus	*Quercus s.s.*	*Quercus s.s.*	4
Total	***Quercus s.s.***	***Quercus s.s.***	**42**
Aphelonyx	*Cerris*	*Cerris*	2
Dryocosmus	*Cerris*	*Cerris*	7
Plagiotrochus	*Cerris*	*Cerris*	14
Pseudoneuroterus	*Cerris*	*Cerris*	1
Neuroterus	*Cerris*	*Cerris*	6
Synophrus†	*Cerris*	*Cerris*	3
Total	***Cerris***	***Cerris***	**30**
Andricus	*Cerris*	*Quercus s.s.*	55
Callirhytis	*Quercus s.s.*	*Cerris*	7

† tribe Synergini

1.4.6 Host alternation

Western Palaearctic oak gallwasp host-alternation almost always involves strict alternation of generations between oaks in the sections *Cerris* and *Quercus s.s.* (Folliot 1964; Askew 1984; Nieves-Aldrey 1992; Cook *et al.* 2002; Stone *et al.* 2001, 2002b, Ács *et al.* 2007). The native distribution of such host-alternating gallwasps is limited to a latitudinal band in which oaks from both host sections co-occur (Section 1.3.3). North of this band, the distributions of host-alternating gallwasps are constrained by the absence of section *Cerris* oaks, while south of this band they are constrained by the absence of section *Quercus s.s.* oaks.

Host-alternating *Andricus* species belong to a single lineage whose ancestral life cycle involved only oaks in section *Quercus s.s.* (Cook *et al.* 2002). The sexual generation galls always develop on section *Cerris* oaks, and the parthenogenetic generation galls always develop on section *Quercus* oaks. At least ten members of this clade have extremely wide longitudinal ranges, extending from Iberia and Morocco to Iran and the Caucasus, and with populations in all of the major European oak refugia (the Iberian Peninsula, Italy, and the Balkans) (Atkinson 2000; Melika *et al.* 2000; Nieves-Aldrey 2001; Cook *et al.* 2002; Rokas *et al.* 2003a,b). A feature of these widespread gallwasps is that while their parthenogenetic generation oak hosts are found over much of this longitudinal range, the sexual generation oak changes with longitude, from *Q. suber* in the Iberian Peninsula and northwestern Africa, to *Q. cerris* from Italy eastwards to Turkey and the Levant, and then to *Q. libani* and *Q. brantii* (Lebanese and Persian oaks) in Iran (Section 1.3.4). Both generations of this life cycle are formally known for only 10 of the 55 *Andricus* species in the host-altrernating clade. A host-alternating life cycle is inferred for the remaining species by DNA sequence data (Cook *et al.* 2002), and by ongoing work confirming the existence of previously unmatched generations on the appropriate oaks (Cook *et al.* 2002; Walker 2002; Rokas *et al.* 2003b; Ács *et al.* 2007). In *Callirhytis* the direction of alternation is usually reversed (Nieves-Aldrey 1992), with sexual generations primarily on oaks in section *Quercus s.s.* and parthenogenetic generations galling oaks in section *Cerris*.

1.4.7 Geographic origins of the oak gallwasps

No phylogenetic analysis of gallwasp relationships across the Holarctic has been completed so identification of centres of origin and dispersal routes of the oak gallwasps to date has been based on identification of centres of diversity and biogeographic separation of extant lineages. Kinsey (1930, 1936) argued that at least some of the oak gallwasp genera, and particularly those allied to *Cynips*, diversified in Mexico and then dispersed to the Palaearctic via the USA and Canada (Kinsey 1930, 1936, 1937; Ritchie & Shorthouse 1987). This pattern is paralleled within the

Nearctic by their section *Quercus s.s.* oak hosts (Manos *et al.* 1999; Manos & Stanford 2001).

While tribal origins remain uncertain, ancient patterns of dispersal and centres of origin have been examined within the Western Palaearctic using phylogenetic and/or population genetic approaches for widespread host-alternating *Andricus* species (*A. dentimitratus, A. grossulariae, A. kollari* and *A. quercustozae*), *Biorhiza pallida* and *Cynips quercusfolii* (Atkinson 2000; Rokas *et al.* 2001; Stone *et al.* 2001, 2002; Rokas *et al.* 2003a; Hayward & Stone 2006). Levels of intraspecific sequence divergence within each of these species, except *C. quercusfolii*, suggest the most recent common ancestor (MRCA) of each species existed several million years ago. Populations of these species in each of Iran, Turkey and the southern European glacial refugia are also genetically distinct, suggesting that they were isolated from each other for substantial periods of time. This isolation is probably due to disruption of the continuity of oak distributions, and hence of gallwasp habitat, during periods of glacial advance. The two main genetic divides within these widespread species are between the Iberian Peninsula and the rest of Europe, and between Europe and Turkey/Iran. There is only limited differentiation between populations in Italy and the Balkans, suggesting genetic exchange either across the Adriatic Sea or south of the Alps (Stone & Sunnucks 1993; Rokas *et al.* 2003a).

A single ancient host shift from *Q. cerris* to *Q. suber* during colonisation of the Iberian Peninsula is inferred for *A. grossulariae, A. kollari* (the switch resulted in the origin of a sibling lineage now recognised as a distinct species, *A. hispanicus*) and *A. quercustozae*. The MRCA of *A. hispanicus* is estimated to have existed between 200,000 – 400,000 years ago (Hayward & Stone 2006), relatively recent compared with MRCAs of Eastern clades (780,000 – 1,900,000 years ago). Combined with declining gradients of allozyme diversity from east to west (Stone *et al.* 2001; Rokas *et al.* 2003a) this pattern has lead to the 'Out of Anatolia' (Rokas *et al.* 2003a) hypothesis for gallwasp origins, and parallels the pattern seen in proposed

centres of origin and taxonomic diversity of oaks (Atkinson *et al.* 2006; Section 1.3.1).

1.4.8 Human impacts on oak gallwasp distributions

Planting of *Q. cerris* (Section 1.3.5) within a natural matrix of native *Q. petraea* and *Q. robur* has allowed at least 11 oak gallwasps to invade northwestern Europe. Nine host-alternating *Andricus* species (*A. aries, A. corruptrix, A. gemmeus, A. grossulariae, A. kollari, A. lignicolus, A. lucidus, A. quercuscalicis* and *A. quercustozae*), and two species currently thought to be wholly dependent on *Q. cerris* (*Aphelonyx cerricola* and *Neuroterus saliens*), have expanded their distributions northwards and all except two (*A. gemmeus, A. quercustozae*) have reached Britain (Claridge 1962; Folliot 1964; Quinlan 1974; Stone & Sunnucks 1992, 1993; Schönrogge *et al.* 1994, 1998; Dauphin 1996, 1999; Crawley 1997; Dauphin & Aniotsbeherre 1997; Csóka *et al.* 1998; Walker 2001; Wurzell 2000; Kwast 2001; Stone *et al.* 2001, 2002a, b; Walker *et al.* 2002). *Andricus quercustozae* is found as far north as southern Brittany in France, and *A. gemmeus* has reached southern Germany (Kwast 2001). It is likely that the distributions of other gallwasps that have a sexual generation on section *Cerris* have also been affected by *Q. cerris*. The expected phylogeographic signature of such range expansion would be populations in the invaded range having a subset of the genetic diversity found in adjacent refugia, which is the pattern that has been observed in *Andricus quercuscalicis* (Stone & Sunnucks 1993) and *Andricus kollari* (Stone *et al.* 2001).

Oak cynipid galls contain high levels of tannins, and have been traded in the Mediterranean region for at least two and a half thousand years, since Theophrastus (371-286 B.C.) recorded the use of species including *A. kollari* in medicine, tanning of leather and the manufacture of inks in his *Historia Plantarum* (Beavis 1988). The scale of this trade is illustrated by Niebuhr's (1776-1780) account of a caravan of 1,000 camels exporting an estimated 10 million oak galls (many of which will have still contained gallwasps) westwards from Syria. Parthenogenetic generation galls of *A. kollari* were imported in large numbers to southern England in the 1830s as a

source of tannins for the manufacture of dyes (Walker *et al.* 2002), initially in Devon, and later in southeast England. The presence of introduced *Q. cerris* alongside extensive plantations of *Q. robur* and *Q. petraea* allowed *A. kollari* to spread unassisted as far as Scotland within 40 years of its introduction. *A. kollari* was the first host-alternating gallwasp to reach Britain, and is now the most widespread host-alternating invader in northern Europe. Trade on this scale is expected to leave a distinct phylogeographic signature with introduced populations having a subset of the genetic diversity of the source population rather than of the adjacent native populations.

1.5 Oak cynipid inquilines

A fundamental requirement of longitudinal phylogeography is reliable taxonomy and distribution information so each species can be sampled across its entire range. Despite their close association with the relatively well-understood oak gallwasps, the taxonomy of the oak cynipid inquilines is still uncertain. The oak inquilines therefore provide an excellent group in which to explore issues in DNA taxonomy and phylogeny reconstruction to provide a foundation for future phylogeographic studies of an important component of oak gallwasp communities. Current understanding of oak inquiline taxonomy and biology are outlined below.

1.5.1 Introduction to the oak inquilines

The tribe Synergini includes all of the Cynipid gallwasps known to develop as inquilines within galls induced by other Cynipids. The evolutionary origins of the Synergini are unclear. Analyses based on adult morphology (Ronquist 1994; Liljeblad & Ronquist 1998) have suggested that the Synergini form a monophyletic group, derived from the Aylacini. Recent DNA sequence-based analysis (Nylander *et al.* 2004a), however, suggests that the Synergini, as currently defined, are polyphyletic within the Cynipinae. According to the current taxonomy, the tribe Synergini includes eight genera, which are predominantly restricted to the temperate zones of the Holarctic region. The most species rich genus, *Synergus*, has

historically been divided into two sections (Sections I and II; Mayr 1872). These sections were proposed on the basis of morphology, however they are also associated with biological differences. Section I species are univoltine (a single generation per year; Section 1.5.3), and their development in a host gall is rarely lethal to the gall-inducer (Section 1.5.5). In contrast, Section II species are bivoltine (two generations per year; Section 1.5.3), and their attack frequently causes death of the gall inducer (Section 1.5.5).

1.5.2 Western Palaearctic inquiline diversity

There are 38 Western Palaearctic species in the tribe Synergini (Pujade-Villar *et al.* 2003). Although the Western Palaearctic inquiline fauna is relatively well known (Pujade-Villar *et al.* 2003), assessment of regional patterns of diversity in the Synergini is affected by the bias in sampling effort across the region. Inquiline species richness, in contrast to patterns seen among the Cynipini (Section 1.4.2), is very low to the east of Europe, almost certainly due to lower sampling effort. Areas such as North Africa, Israel, Iran, and the Transcaucasus have only been sampled opportunistically (Nieves-Aldrey *et al.* 2007). Based on current knowledge of species distributions, the 38 Western Palaearctic Synergini can be divided into four biogeographical distributions: (i) species found across the entire range of Western Palaearctic oaks (17 species); (ii) species found across southern parts of the Western Palaearctic (8); (iii) southern species absent from the Iberian Peninsula and northwestern Africa (6); and (iv) species found only in the Iberian Peninsula and northwestern Africa (5).

1.5.3 Oak inquiline life cycles

Most of the 38 Western Palaearctic Cynipid inquiline species have similar life cycles, with one or two generations per year, synchronised with their hosts. Adults usually emerge from the galls after the gall inducers and lay eggs in freshly initiated galls. In bivoltine *Synergus* species, adult morphology may differ markedly between generations (Wiebes-Rijks 1979; Pujade-Villar 1992a). Adult size can also vary

dramatically, and appears to be directly related to the size of the larval chamber (Nieves-Aldrey *et al.* 2007). In *Synergus gallaepomiformis*, the morphological differences between generations are under environmental control (Wiebes-Rijks 1979).

1.5.4 Host associations

The prevalence of broad host associations is likely to increase as species are sampled across a wider range of sites. The prevalence of narrow host associations is also likely to increase with the discovery of new species in areas of probable high diversity. Thirty-three of the 38 Western Palaearctic inquiline species are associated with galls on widespread oaks and are typically found on only one of the oak sections (Nieves-Aldrey *et al.* 2007). Seventeen species are associated with oaks in section *Quercus s.s.* and 14 are associated with oaks in section *Cerris*. Only two species are found on oaks in both sections.

Seventeen of the Western Palaearctic inquiline species are associated with a single host genus, including nine with *Andricus*, two with *Aphelonyx* and four with *Plagiotrochus*. In contrast, six species are present in galls of five or more host genera. The number of inquiline species associated with each host cynipid genus increases with the number of host gall forms per genus, with the exception of *Neuroterus*, which has more inquiline species associated with its galls than expected (Nieves-Aldrey *et al.* 2007).

1.5.5 Effects of inquilinism on the host gallwasp and gall

Inquilines can be classified as lethal (always cause the death of the host gallwasp), non-lethal (never cause the death of the host) and facultative (the fate of the host depends on the location of the inquilines in the gall). Inquiline larvae never consume the host gallwasp but feed entirely on gall tissue. Death of the host results either if the larva is stung by the ovipositing female or if it is crushed by the developing inquiline larva (Evans 1967, 1972; Weld 1952). Lethal cynipid inquilines can be

major causes of gallwasp mortality. *Synergus pallipes* and *Synergus clandestinus*, for example, cause significant host mortality in parthenogenetic generation galls of *Cynips divisa* (Sitch *et al.* 1988) and *Andricus legitimus* (Wiebes-Rijks 1980), respectively.

Attack by lethal inquilines often results in cessation of gall growth, and the development of a distorted and stunted gall (Evans 1967). In contrast, the presence of non-lethal inquilines may cause increased gall tissue formation (Brooks & Shorthouse 1998). The most dramatic modification of oak gall structure by a Cynipid inquiline is that induced by members of the genus *Synophrus*. *Synophrus politus* attacks sexual generation Cynipid galls induced by species in the *Andricus burgundus* complex at an early stage in development (Pujade-Villar *et al.* 2003). The original tissues of the gall are rapidly reabsorbed and the resulting gall, while initially irregularly-shaped and soft, becomes approximately spherical and among the hardest of all oak galls (Pujade-Villar *et al.* 2003).

1.6 Thesis outline

Chapter 2 – Phylogenetic methods.

In order to address the aims of this thesis, it is important to select appropriate phylogenetic methods. This chapter discusses several important issues, such as phylogenetic techniques, the role of models and model selection and some specific considerations of Bayesian phylogenetics, to provide a justification for the methods adopted in subsequent chapters. Specific details of molecular methods and bioinformatic software used in this thesis are also provided.

Chapter 3 – Longitudinal phylogeography of the Western Palaearctic: Patterns and methodology.

The traditional approach to the phylogeography of the Western Palaearctic has been to assess latitudinal patterns of post-glacial colonisation. This approach treats refugial populations as sources of diversity, thereby ignoring underlying longitudinal patterns. A historic sampling bias has led to many studies ignoring populations in

the east of the region. Longitudinal phylogeography requires relatively even sampling across a species entire range in order to determine species' origins and directions of range expansion. This chapter presents a review of the Western Palaearctic phylogeographic literature to determine the current status of longitudinal phylogeography and to address the following questions: What is the extent of sampling bias in the Western Palaearctic? Do studies in which a species distribution has been evenly sampled provide evidence of a geographic origin? Are there shared directions and timescales of longitudinal range expansion across taxa?

Chapter 4 – Longitudinal range expansion and cryptic eastern species in the Western Palaearctic oak gallwasp *Andricus coriarius*.

Andricus coriarius is a widespread Western Palaearctic oak gallwasp, distributed from Iberia to Iran and with populations in each of the major oak refugia. Unlike many other host-alternating oak gallwasp species, the phylogeographic distribution of *A. coriarius* is expected to have been relatively unaffected by human activity, since it has not been directly traded and has not invaded northern Europe following the introduction of *Q. cerris*. This chapter presents the first phylogeographic analysis across the entire longitudinal range of a widespread oak gallwasp to address the following questions: What is the extent of genetic diversity in areas to the east of Europe? Is the 'out of Anatolia' hypothesis of eastern origin for Western Palaearctic oak gallwasps supported? The work presented in this chapter has resulted in two publications: one in *Molecular Ecology* on phylogeographic structure (Challis *et al.* 2007), and one in *Zootaxa* describing new species identified as a result of this work (Melika *et al.* 2007).

Chapter 5 – The phylogeographic clade trade: Tracing the impact of human-mediated dispersal on the colonisation of Northern Europe by the oak gallwasp *Andricus kollari*.

Andricus kollari has been the most widely traded oak gallwasp. Although the history of deliberate dispersal of *A. kollari*, first across the Mediterranean and later into the UK, is well documented, the origin of extant populations is unknown. This chapter presents a phylogeographic analysis of *A. kollari* to address the following questions:

Does the legacy of trade in the Mediterranean prevent phylogeographic reconstruction from a longitudinal perspective? Which geographic region contained the source population for colonisation of the UK? Has the introduced UK population provided a source for colonisation of the rest of northern Europe? The work presented in this chapter has resulted in a paper in *Molecular Ecology* (Stone *et al.* 2007c).

Chapter 6 – Evidence for genetic introgression in the phylogeography of the *Andricus quercuscalicis* clade.

Uniquely among insects, closely related gallwasp species are believed to have undergone intrarefugial introgression. Intrarefugial introgression presents a problem for longitudinal phylogeography as it violates the assumption that molecular markers within refugia can be used to infer the phylogeographic history of a single species. Within the oak gallwasps, this genetic exchange has been inferred to be most prevalent among members of the host-alternating *Andricus quercuscalicis* clade. This clade includes *Andricus quercustozae/insana*, which show marked regional variation in gall morphology and have been considered as separate species. This chapter presents a DNA barcoding and phylogeographic analysis of the *Andricus quercuscalicis* clade to address the following questions: Can DNA barcoding be used to resolve species-level clusters in light of introgression? Do sequences within the clade predominantly cluster by species or by geography? Are geographic divisions concordant with those reported for other oak gallwasp species? Are regional speciation hypotheses (e.g. *A. quercustozae/insana*) supported?

Chapter 7 – Comparative phylogeography of Western Palaearctic oak gallwasps.

Longitudinal phylogeography has the potential to reveal shared patterns of origin and range expansion across species, which may have important implications for setting conservation priorities. Oaks support diverse invertebrate communities and the distributions of oak-associated invertebrate species are determined by the distribution of oaks. Longitudinal concordance among oak-associated invertebrate species may

be expected if their origins and patterns of range expansion are also determined by those of the oaks. The oak gallwasps are an excellent group in which to test this hypothesis as they are a large group of species with a particularly close association with their oak hosts. This chapter presents a comparative phylogeographic analysis of host-alternating and non-alternating oak gallwasps to address the following questions: Do the oak gallwasps share an eastern origin with their oak hosts? Do the oak gallwasps have shared directions and timescales of range expansion? Have the different host requirements affected phylogeographic concordance between host-alternators and non-alternators?

Chapter 8 – Phylogenetic relationships among Western Palaearctic species of the inquiline gallwasp tribe Synergini (Hymenoptera, Cynipidae): a test of molecular barcoding and an examination of the trade-off between data quality and maximal taxon sampling.

Oak cynipid inquilines are important components of oak gallwasp communities yet they are relatively poorly understood. The separation of closely related species presents a challenge to morphological taxonomy since the degree of variation across a species range is often greater than that between separate species. Appropriate species delimitation and inference of relationships among closely related species will be essential in understanding the evolution of traits, such as host specificity and lethality, which affect oak gallwasp community composition and development. Species delimitation is also a vital first step towards longitudinal phylogeographic analysis of the oak inquilines as it will allow the entire ranges of true species to be sampled for assessment of concordance with the oak gallwasps. This chapter presents DNA barcoding and multigene phylogenetic analyses of the oak inquilines to address the following questions: Are molecular taxa defined using DNA barcodes congruent with morphological taxonomy? Are DNA barcodes consistent across different nuclear and mitochondrial genes? Are divisions into genera supported by a multigene phylogeny? Are Mayr's sections supported? How does incorporating missing data into a relatively small (three gene) alignment to improve taxon sampling affect the accuracy of phylogenetic reconstruction?

Chapter 9 – Concluding remarks.
This chapter provides a summary of the main findings of this thesis and identifies further questions to be addressed.

1.7 Contributions to the work in this thesis

1.7.1 Sample collection

Samples for use in this thesis were collected over a period of 10 years, both specifically for the work presented in this thesis and during sample collection trips for related projects. Richard Challis collected samples from Hungary, Spain and southwest France for use in Chapter 7. Additional samples for Chapters 4 to 8 were collected by Zoltan Ács, Alex Aebi, Rachel Atkinson, Tracey Begg, Gordon Brown, Amy Crowther, György Csóka, Alex Hayward, Antonio Hernandez-Lopez, Gil McVean, George Melika, Serap Mutun, James Nichols, Jose-Luis Nieves-Aldrey, Rida Nuwayid, Olivier Plantard, Sonja Preuss, Juli Pujade-Villar, Antonis Rokas, Alex Rowe, Ebrahim Sadeghi, Karsten Schönrogge, Graham Stone and Majid Tavakoli.

1.7.2 Laboratory work

Sequence data for Chapters 4 to 7 were generated by Alex Aebi, Alex Hayward, Antonio Hernandez-Lopez, James Nichols, Sonja Preuss, Antonis Rokas and Graham Stone. Sequence data for Chapter 8 were generated by Zoltan Ács and Alex Hayward. Richard Challis generated cytochrome *b* sequence data for *Cynips quercusfolii* for Chapter 7 and some sequence data for the D2 region of the nuclear ribosomal 28S gene for Chapter 5. Chapters 4 to 8 also make use of previously published sequence data.

Allozyme data for Chapter 5 were generated by Graham Stone and Rachel Atkinson. Chapter 5 also makes use of previously published allozyme data.

1.7.3 Data analysis

Richard Challis performed all analyses of sequence data for this thesis and the Structure (Section 2.12.14) analysis of allozyme data for Chapter 5. The remaining allozyme data analyses in Chapter 5 were performed by Graham Stone.

1.7.4 Writing

All chapters were principally written by Richard Challis with the exception of Chapter 5, which is based on a published manuscript to which both Richard Challis and Graham Stone contributed equally. All chapters benefited from extensive comments from Amy Crowther and Graham Stone. Specific chapters also benefited from comments and contributions from Zoltan Ács (Chapter 8), Mark Blaxter (Chapter 8), Richard Ennos (Chapter 3), George Melika (Chapters 6 and 8), James Nichols (Chapter 7), Zsolt Pénzes (Chapter 8) and Andrew Rambaut (Chapter 7).

Chapter 2

Phylogenetic methodology

Chapter 1 introduced the study system and gave an overview of the aims of this thesis. This chapter provides a critical review of phylogenetic methods and describes the methodological approaches adopted in this thesis. Part 1 reviews current methods for phylogenetic reconstruction and analysis of molecular sequence data to provide a background to, and justification for, the methods adopted in this thesis. Part 2 describes the protocols and software for the molecular and bioinformatic methods that are used in subsequent chapters. Specific details of methods used are presented in each chapter.

Part 1 – Review of phylogenetic methodology

2.1 Molecular markers

Molecular markers are genotypic characters that can provide evidence of the relationships among individuals, populations and species. Although only three classes of molecule are typically used (DNA, RNA or protein), a large number of molecular marker approaches have been developed to distinguish between individuals according to: size, shape or charge of molecules or fragments of molecules; chromosomal banding patterns; molecular sequence (especially for DNA); and genome-scale features such as gene order and even changes in the genetic code. DNA sequences and allozymes (enzymes produced by different alleles of the same gene) are discussed below to illustrate some of the differing properties of different classes of molecular marker.

2.1.1 DNA sequences

DNA sequences provide the most informative molecular markers and have become the most widely used markers for molecular phylogenetics. The redundancy of the genetic code and regions of non-coding DNA mean that many genetic differences between individuals can only be observed at the DNA sequence level. The processes

underlying these differences can be modelled (Section 2.3) and, provided an outgroup is available, the phylogenetic relationships between sets of DNA sequences can be rooted. Rooted phylogenies provide evidence about which lineages are basal and which are derived and trait-mapping (Section 2.9) allows the most likely character-state of the basal node to be inferred (e.g. Allan *et al.* 2004).

The widespread use of DNA sequences as molecular markers has been made possible through the development of the polymerase chain reaction (PCR). For many genes, universal primers have been developed to allow PCR amplification of DNA for the same gene from a range of taxa so sequences for new samples can be obtained relatively rapidly. One drawback to the use of DNA sequences is that patterns for a single marker may not represent the phylogenetic history of a species. In particular there is often discrepancy between the phylogenetic history of nuclear and organellar DNA (Hurst & Jiggins 2005) so unless multiple DNA sequence markers are used the resulting analysis will reflect the history of the gene that has been sequenced, rather than the species.

2.1.2 Allozymes

Allozymes variation may be distinguished by the different physical and chemical properties of different enzyme alleles. Changes in the size and charge of a protein affect the rate at which it moves through a gel so different alleles can be distinguished. This technique is conservative as many DNA sequence differences will not affect the amino acid composition of the protein and some amino acid changes will not affect the rate at which the allozyme moves through a gel. The other drawback to using markers such as allozymes is that they are unrootable. Unrootable markers do not provide information on the relative age of clades. Clades in unrooted trees should strictly be referred to as 'clans' (Wilkinson *et al.* 2007) since uncertainty over the placement of a root means that members of a clan cannot be defined as sharing an common ancestor that is not shared by any other taxon. Geographic origin may be inferred on the assumption that the centre of highest diversity is likely to represent the origin, building upon the principle that the centres

of diversity for any marker represent refugia. Stone and Sunnucks (1993) illustrate the pattern of diversity at putative refugia for a historically documented range expansion. The pattern is well documented in plants (Broyles 1998, Lagerkrantz & Ryman 1990, Demesure *et al.* 1996, Dumolin Lapegue *et al.* 1997, King & Ferris 1998, Tomaru *et al.* 1997), insects (Armbruster *et al.* 1998, Hard *et al.* 1993, Cooper *et al.* 1995, Wilcock *et al.* 2001) and vertebrates (Ball *et al.* 1988, Merilä *et al.* 1997, Sage & Wolff 1986). Increased sampling intensity for the same markers used in an unresolved study may allow inference of geographic origin. Such patterns of diversity should be interpreted with caution, however, as derived populations may have high diversity due to admixture from multiple sources (Fry & Zink 1998, Green *et al.* 1996, Zheng & Ennos 1999), and allelic diversity and heterozygosity can give conflicting signals (Comps *et al.* 2001).

2.1.3 Molecular markers for this thesis

Where possible DNA sequences will be obtained for at least one nuclear and one mitochondrial marker to reconstruct phylogenetic and phylogeographic relationships in this thesis. Allozyme data will also be used in conjunction with DNA sequence data to provide additional resolution.

2.2 Techniques

Four major techniques have been proposed for phylogenetic reconstruction: parsimony, neighbour-joining, maximum likelihood and Bayesian inference. In order to select the most appropriate methods for this thesis, it is important to have an understanding of the underlying assumptions and computational requirements of each of these techniques.

2.2.1 Parsimony

Parsimony, **MP**, is the oldest and perhaps the most intuitive method for phylogenetic reconstruction, appealing to the philosophical principal that the correct answer is usually the one requiring the fewest assumptions. MP was first applied to

phylogenetic analysis by Edwards & Cavalli-Sforza (1964). While MP remains an important method, it has been shown to be statistically inconsistent (Felsenstein 1978b). The validity of the assumption that the simplest hypothesis is true has been brought into further doubt by studies revealing that complex structures, which MP predicts would evolve only once, have evolved on multiple occasions (Collin & Cipriani 2003; Whiting *et al.* 2003).

2.2.2 Neighbour-joining

Neighbour-joining, NJ (Saitou & Nei 1987), begins with a star-like phylogeny of all taxa and progressively clusters pairs of taxa that minimise the total branch length. Although NJ is probably the least computationally intensive of the commonly used methods for phylogenetic reconstruction, it is flexible. The ability to use model based distance estimates (Section 2.3) and produce a phylogeny with support indices (Section 2.6) in a relatively short time have promoted the continued use of NJ despite the development of more sophisticated methods.

2.2.3 Maximum likelihood

Unlike MP and NJ, **maximum likelihood, ML**, attempts to determine the hypothesis most likely to explain the data, under a specific model of evolution (although NJ can use model-based distance estimates). Formally applied to DNA sequence evolution by Felsenstein (1981), this model-based approach benefits from the capacity to apply different rates to, for example, nucleotide transitions and transversions (Section 2.3.1.1). The drawback to model-based approaches are two-fold: (i) the use of an inappropriate evolutionary model may detrimentally affect analysis (Farris, 1999); and (ii) the analysis of large datasets, particularly under complex models is computationally intensive. Evolutionary models can be chosen using model selection criteria (Section 2.3.3).

2.2.4 Bayesian inference

More recently, **Markov chain Monte Carlo, MCMC**, simulation has been used to apply **Bayesian inference, BI**, to phylogenetic reconstruction (Yang & Rannala 1997). BI uses posterior probabilities, i.e. the probability of the phylogeny given the data, to select the most probable topologies. The posterior probability of the i^{th} phylogenetic hypothesis, τ, being correct given data, D, is given by Bayes' rule:

$$P(\tau_i \mid D) = P(D \mid \tau_i)\, P(\tau_i) \frac{P(D \mid \tau_i)\, P(\tau_i)}{\sum_{j=1}^{B(n)} P(D \mid \tau_j)\, P(\tau_j)}$$

Where $P(D|\tau_i)$ is the likelihood function and $P(\tau_i)$ is the prior probability of the i^{th} phylogenetic hypothesis. The posterior probability is obtained by summation over all possible phylogenetic hypotheses, $B(s)$, that are possible for s terminal taxa (integration if s is large).

This cannot be calculated analytically for large datasets, so MCMC techniques were introduced to allow sampling from the posterior probability distribution. The frequency of occurrence of a clade in a MCMC analysis is directly proportional to the posterior probability (Section 2.6) of the clade (Tierney 1994). For each iteration of the Markov chain: (i) model parameters are updated; (ii) a tree is produced using these updated parameters; and (iii) the likelihood of the tree is compared with the likelihood of the tree from the previous iteration. Whereas in ML, the new tree will be retained only if it is more likely than the previous tree, in BI there is a probability that a less likely tree will be retained. This allows likelihood to decrease as well as increase and the Markov chain will eventually converge near the most likely phylogeny. The distribution of phylogenies sampled after convergence provides an approximation of the posterior probability distribution. As with ML, it is possible for a single Markov chain to become trapped at a local maximum. The use of **Metropolis coupled MCMC, $(MC)^3$** (Metropolis *et al.* 1953; Huelsenbeck & Ronquist 2001; Ronquist & Huelsenbeck 2003), addresses this issue through the use of simultaneous Markov chains. The chains rapidly diverge due to the stochasticity

of MCMC and at each iteration there is a probability that two chains will exchange parameters, allowing the analysis to escape local optima.

2.2.5 Techniques for this thesis

BI allows simultaneous estimation of support for an inferred phylogeny (Section 2.6). This technique also allows for the use of a model of sequence evolution (Section 2.3) and accounts for the uncertainty of phylogenetic inference by sampling from a set of most likely topologies, rather than determining a single most likely topology, as for ML. However, the benefits of this technique must be weighed against the high computational intensity and the requirement for explicit description of an evolutionary model (Section 2.3). Explicit prior assumptions, which can detrimentally affect the analysis if incorrectly specified, about the distribution of parameter values (Section 2.4) must also be made. Further, BI cannot be applied to phylogenetic network reconstruction (Section 2.7). BI will therefore be the preferred method for phylogenetic tree reconstruction but the remaining techniques will also be considered for other analyses.

2.3 Modelling sequence evolution

Models of sequence evolution must be specified when using model-based techniques such as ML (Section 2.2.3) and BI (Section 2.2.4). These models attempt to describe patterns of nucleotide substitution for a dataset as accurately as possible while minimising computational intensity. This section describes the parameters of sequence evolution that are commonly modelled and discusses techniques used to select the most appropriate model from a set of possible models.

2.3.1 Ways to model sequence evolution

There are a number of different ways to model sequence evolution. In general, three properties of sequence evolution must be considered: (i) the rate of fixation of nucleotide substitutions; (ii) the rate of fixation of insertions and deletions; and (iii) rate heterogeneity.

2.3.1.1 The rate of fixation of nucleotide substitutions

In the simplest model of evolution, the Jukes-Cantor (JC) model (Jukes & Cantor 1969), all bases within a sequence are permitted to evolve independently but with the same mutation rate (μ). Many more complex models have been proposed in an attempt to provide a more realistic description of sequence evolution. The most commonly applied (Figure 2.1) are: (i) the Kimura two-parameter (K2P) model (Kimura 1980), which allows different rates for transitions (r_{ti}) and transversions (r_{tv}); (ii) the HKY model (Hasegawa, Kishino & Yano, 1985), which is an extension of the K2P in which base frequencies (π_A, π_C, π_G and π_T) are not assumed to be equal; and (iii) the general time reversible (GTR) model (Tavaré 1986) in which base frequencies are not assumed to be equal and transitions or transversions between each pair of nucleotides have independent rates.

K2P:

$$Q = \begin{bmatrix} & [A] & [C] & [G] & [T] \\ [A] & - & r_{tv} & r_{ti} & r_{tv} \\ [C] & r_{tv} & - & r_{tv} & r_{ti} \\ [G] & r_{ti} & r_{tv} & - & r_{tv} \\ [T] & r_{tv} & r_{ti} & r_{tv} & - \end{bmatrix}$$

HKY:

$$Q = \begin{bmatrix} - & \pi_C r_{tv} & \pi_G r_{ti} & \pi_T r_{tv} \\ \pi_A r_{tv} & - & \pi_G r_{tv} & \pi_T r_{ti} \\ \pi_A r_{ti} & \pi_C r_{tv} & - & \pi_T r_{tv} \\ \pi_A r_{tv} & \pi_C r_{ti} & \pi_G r_{tv} & - \end{bmatrix}$$

GTR:

$$Q = \begin{bmatrix} - & \pi_C r_{AC} & \pi_G r_{AG} & \pi_T r_{AT} \\ \pi_A r_{AC} & - & \pi_G r_{CG} & \pi_T r_{CT} \\ \pi_A r_{AG} & \pi_C r_{CG} & - & \pi_T r_{GT} \\ \pi_A r_{AT} & \pi_C r_{CT} & \pi_G r_{GT} & - \end{bmatrix}$$

Figure 2.1 Substitution rate matrices for three commonly applied models of evolution.

2.3.1.2 The rate of fixation of insertions and deletions

Insertions and deletions (indels) result in gaps in sequence alignments. Beyond the ability to code gaps as a fifth character state, the utility of gaps in determining phylogenetic position has been largely ignored. With the advent of phylogenomics,

models are being developed to account for non-sequential differences between taxa, so the use of gaps should become more prominent. However, concerns have been raised about the circularity of using sequence data aligned on a guide tree to generate phylogenies.

2.3.1.3 Rate heterogeneity

Allowing for variation in substitution rates among sites in sequence alignments can dramatically improve the accuracy of a phylogenetic hypothesis. The most common model for rate distribution along nucleotide sequences is the gamma (Γ) distribution (more precisely the probability density function of the gamma distribution):

$$f(r;\alpha,\beta) = r^{\alpha-1}\frac{e^{-r/\beta}}{\beta^{\alpha}\Gamma(\alpha)} \quad for\ r > 0$$

Where the distribution is determined by the shape parameter, α, and the scale parameter, β.

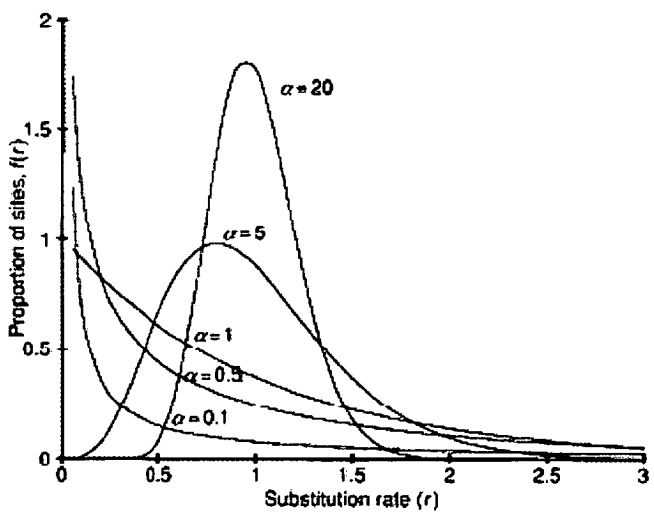


Figure 2.2 The density function of the gamma distribution of substitution rates, taken from Yang (1996).

The gamma distribution has a mean (substitution rate, r) α/β and variance α/β^2. This model is widely used because it is flexible and yet requires estimation of

only a single parameter. Since r for each nucleotide is relative to the rest of the sequence, β is fixed equal to α to give a mean substitution rate of 1. The value of α determines the shape of the distribution (Figure 2.2), L-shaped for $\alpha \leq 1$ (most sites have low substitution rate but some hotspots exist with very high substitution rate) and bell-shaped for $\alpha > 1$ (most sites have intermediate substitution rates). When $\alpha \leq 1$, and the gamma distribution is L-shaped, rate heterogeneity can be described with no further parameters, however, when $\alpha > 1$, and the gamma distribution is bell-shaped, it is often necessary to introduce an extra parameter to account for invariable sites. Computation under the gamma distribution can be simplified through use of a discrete gamma distribution with the density distribution of substitution rates successfully represented by as few as four categories (Yang 1994).

2.3.2 Data partitioning

Data partitioning is an extension of the consideration of rate heterogeneity along a sequence. If different sections within a sequence alignment have evolved independently, then model parameters that represent one section may be inappropriate for another section. Partitioning the data into regions with shared parameters allows the use of a separate model for each data partition, more accurately reflecting the evolutionary history of a sequence alignment. However, the number of model parameters increases approximately linearly with the number of data partitions, so care must be taken to ensure that the partitions are biologically meaningful.

The most commonly applied partitions are divisions between nuclear and mitochondrial genes, coding and non-coding DNA, stem and loop regions of rDNA or codon positions of protein coding genes. Often these partitions are used without testing whether they are appropriate to the specific dataset but the impact of the increased number of parameters should always be assessed (Section 2.3.3). In protein coding sequences, for example, it will only be appropriate to consider the three codon positions separately if they have been subject to different evolutionary pressure. If the three positions are not all independent, it may be more appropriate to

specify a less parameter-rich model by separating only third positions which have greater redundancy than first or second positions or treating the alignment as a single partition. Within a Bayesian framework, the computational penalty of increased parameter richness can be offset by the properties of MCMC simulation (Nylander *et al.* 2004b). Partition congruence can be tested using Bayes factors (Section 2.3.3.4) following the method of Irestedt *et al.* (2004).

2.3.3 Model selection criteria

Model selection methods provide an objective set of criteria with which to determine the most appropriate model of sequence evolution for a dataset. However, different methods may select different models for the same dataset due to differences in the underlying assumptions, which are discussed below.

2.3.3.1 Akaike information criterion

The Kullback-Liebler (K-L) distance (Kullback & Liebler 1951) is a measure of the reduction in likelihood obtained by using an incorrect model in place of the 'true' model. The expected K-L distance can be estimated in phylogenetics by using the **Akaike information criterion, AIC** (Akaike 1974):

$$\mathrm{AIC} = -2\ln L + 2K$$

Where $\ln L$ is the maximised log-likelihood and K is the parameter richness of the model.

As parameter richness is increased, $\ln L$ is expected to decrease while the 'penalty term' increases so the model with the lowest AIC will be a balance between parameter richness and the informativeness of additional parameters.

The AIC can be expressed more usefully in terms of the weight of support for each of a hierarchical series of models. The Akaike weight, w, for the ith model in a set of models R is:

$$w_i = \frac{e^{(-1/2\Delta_i)}}{\sum_{r=1}^{R} e^{(-1/2\Delta_r)}}$$

Cumulative Akaike weights for each of the R models (in order of highest w) can be used to determine the 95% set of models that minimise K-L. This provides a way to quantify model selection uncertainty.

2.3.3.2 Bayesian information criterion

The **Bayesian information criterion, BIC** (Schwartz 1978), approximates marginal ln$L$ for a set of candidate models:

$$\text{BIC} = -2\ln L + K \ln n$$

Where n is the sample size (in phylogenetics, $n \approx$ the number of characters in the sequence alignment).

Since the BIC approximates marginal ln$L$, the difference between two BIC estimates provides an approximation to Bayes factors (Section 2.3.3.4) with considerably less computational effort. A true marginal ln$L$ is calculated by integrating over (marginalising) model parameters to estimate the probability of the data given a model, irrespective of specific parameter values. This value can also be estimated by calculating the harmonic mean of the ln$L$ values from MCMC sampling (Section 2.3.3.4). Since each generation uses different parameter values, the influence of specific parameter values are marginalised when a single harmonic mean is calculated.

2.3.3.3 Likelihood ratio tests

Likelihood ratio tests, LRTs, compare model likelihoods directly. The LRT test statistic for model 1 over the null model 0 is:

$$LRT = 2(\ln L_1 - \ln L_0)$$

If the hypotheses are nested, this is a hierarchical likelihood ratio test (hLRT; Frati *et al.* 1997; Huelsebeck & Crandall 1997), which can be assessed against the χ^2 distribution (Goldman 1993). The significance of non-hierarchical LRTs can also be assessed by using the rule of thumb that two ln*L*s constitutes a significant difference (Edwards 1972; Pagel 1999) to avoid computationally intensive Monte Carlo methods (Goldman 1993).

In phylogenetic model testing, the one-parameter JC69 model can be obtained from the two-parameter K80 model by assuming that transitions and transversions occur at the same rate so JC69 is nested within K80. Further nesting levels are shown in Figure 2.3. LRTs are perceived to be more accurate than the approximate AIC and BIC approaches as they allow comparison of actual maximum likelihoods from analyses under different models. LRTs are also convenient as they can be applied after analyses have been performed. There are, however, problems with the LRT approach (Sanderson & Kim 2000): (i) the non-hierarchical LRT method for non-nested models can only provide an approximate result; (ii) situations exist for which an optimal model cannot be selected by hLRTs; (iii) the starting point can affect the model selection procedure; and (iv) hLRTs involve performing multiple tests with the same data which may lead to false positives.

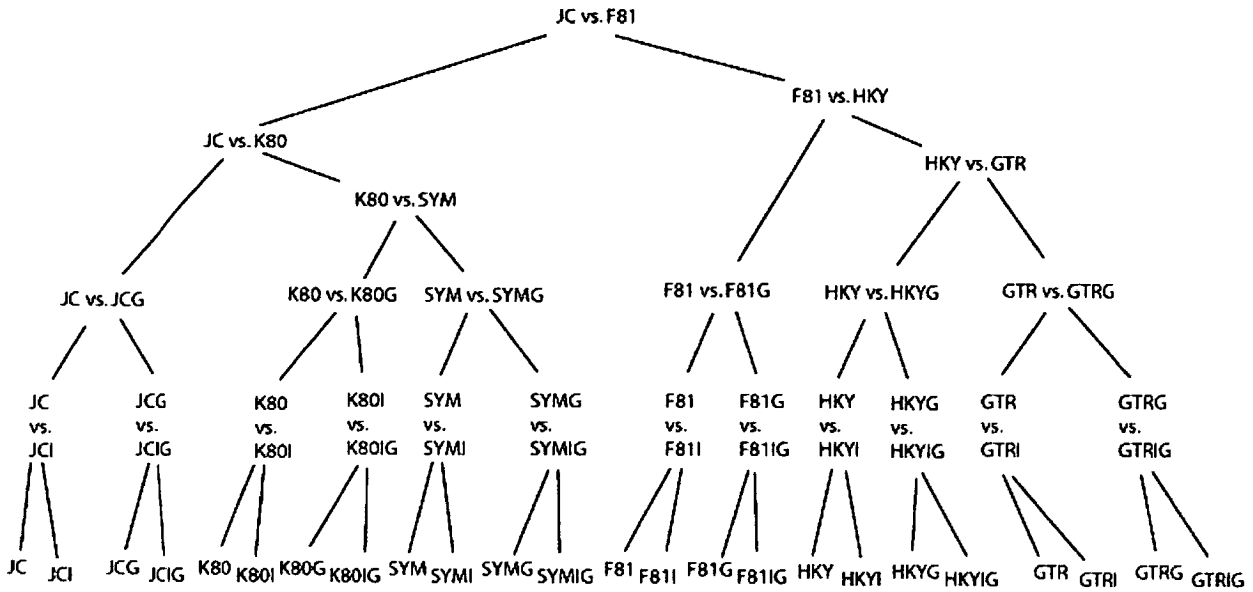


Figure 2.3 Modeltest hLRT hierarchy (after Posada and Crandall 2001).

2.3.3.4 Bayes factors

Bayes factors, BF, were introduced by Kass & Raftery (1995) as a Bayesian equivalent to LRTs. For BF, the marginal likelihoods under each model are considered in place of the likelihoods and the support for the most probable model can be quantified using the criteria of Kass and Raftery (1995). BF have the advantage that, unlike LRTs, the same test can theoretically be applied to both nested and non-nested data. The major drawback is that the 'critical value' for strong evidence of support for *a* model, which is typically simplified to > 10 (e.g. Irestedt *et al.* 2004) should, according to Kass & Raftery (1995) be set at > 20 in phylogenetics to reduce the occurrence of false positives. Results of studies using Bayes factors should, therefore, be interpreted with caution.

2.3.4 Model averaging

Within a Bayesian framework, it is possible to extend the model-testing approach from selection of a single best model to model averaging over the set of candidate models. In practise, this is straightforward as parameter estimates for each candidate

model can be sampled from the stationary phase of a MCMC run in proportion to the probability of that model. Probabilities for each model (or the set of models for which the cumulative probability is at least 95%) can be obtained through AIC or BIC weights (as described above for AIC weights), or through BF by determining the posterior probability for each candidate model. For R candidate models, the posterior probability of the ith model is:

$$P(M_i \mid D) = \frac{P(D \mid M_i)P(M_i)}{\sum_{r=1}^{R} P(D \mid M_r)P(M_r)}$$

2.3.5 Model selection for this thesis

Where phylogenies for this thesis are reconstructed using BI, model averaging will be used in order to account for model selection uncertainty. While the use of posterior probabilities is the 'purest' Bayesian approach to model averaging, uncertainty in prior specification can affect the posterior probability assigned to each model (Section 2.4; this is also an issue in the use of BIC). AIC weights are therefore expected to provide the best guide to model probabilities for use in Bayesian model averaging. For ML and NJ analyses, a single model will be selected using AIC. Selection between alternate partitioning strategies will be performed using BF.

2.4 Choosing and using Bayesian priors

2.4.1 The role of priors

Priors are perhaps the most controversial aspect of Bayesian phylogenetic inference. They represent the knowledge or assumptions, before analysis, about the distribution of probabilities of all possible hypotheses for each parameter and are included as the $P(\tau)$ term in Bayes rule:

$$P(\tau_i \mid D) = P(D \mid \tau_i)\, P(\tau_i) \frac{P(D \mid \tau_i)\, P(\tau_i)}{\sum_{j=1}^{B(n)} P(D \mid \tau_j)\, P(\tau_j)}$$

So a prior is essentially a factor by which the likelihood terms are multiplied to account for the fact that all hypotheses may not be equally likely *a priori*.

2.4.2 Flat priors

In a typical Bayesian analysis with several parameters, each parameter has its own prior. The most commonly applied priors are 'flat' priors, $P(\tau) = 1/B_{(n)}$, in which all hypotheses are considered to be equally likely *a priori*. One restriction on the use of flat priors is that the sum of the probability across all priors must be equal to one, so a flat prior on a continuous parameter must have both upper and lower limits.

In many situations in which it would be possible to set more informative priors, the efficiency of MCMC simulation renders it unnecessary. Nucleotide frequencies, for example, have a significant impact on most models of evolution and can be determined before analysis. Priors should, however, be estimated from independent data to that used in the analysis. Nucleotide frequencies should, ideally, be estimated from the whole genome. In taxa for which a full genome sequence is unavailable, it may be best to set a flat prior on nucleotide frequencies.

Flat priors are often termed uninformative priors and their use is often considered to represent a conservative approach as it can be applied whether or not additional information is available for a parameter in the taxa under investigation. However, the *a priori* assumption that all possible values of a parameter are equally likely actually represents a very strong prior.

2.4.3 Specifying informative priors

There are situations in which a flat prior is inappropriate. Branch lengths, in particular, are frequently modelled with exponential (Ronquist & Huelsenbeck 2003) or coalescent (Drummond *et al.* 2006) priors. Understanding of the biological processes involved in sequence evolution can guide prior selection to improve

phylogenetic inference. However, incorrect prior estimation may adversely affect the outcome of an analysis by delaying or even preventing convergence on the 'correct' hypothesis. Prior sensitivity analysis (e.g. Wilson & McVean 2006) can be applied to allow *a posteriori* testing of the impact of *a priori* assumptions.

2.4.4 Specifying priors for this thesis

Specifying priors requires knowledge of the expected distribution of each model parameter. In most cases, default priors will be accepted in the subsequent chapters in order to reduce the likelihood of misspecification of priors. A notable exception is in the testing of the validity of molecular clock hypotheses (Section X, clocks) in which branch length priors will be altered.

2.5 Diagnosing convergence

A Markov chain that is properly set up and run for an infinite number of iterations should provide a representative sample from the target distribution (Tierney 1994). The most important test of the validity of a Bayesian phylogenetic analysis is thorough post-run analysis to ensure that the MCMC run has converged. Convergence diagnostics fall into two categories: (i) those that indicate whether the stationary phase of the Markov chain has been reached (e.g. split frequencies and parameter value plots); and (ii) those that indicate whether the sample from the stationary phase is representative of the target distribution (e.g. effective sample size, potential scale reduction factor and parameter density plots).

2.5.1 Split frequencies

The most reliable way to determine whether the stationary phase of a MCMC run has been reached is to perform two independent runs (from different initial topologies) and measure the difference in likelihood estimates from the two chains for each iteration. Two runs can be assumed to have converged on the same stationary distribution if the standard deviation of split frequencies is less than 0.01 (Huelsenbeck & Ronquist 2003).

2.5.2 Parameter value plots

For a single run, the simplest way to assess convergence is to plot log-likelihood values against iteration number. Once log-likelihood values stop increasing and begin to fluctuate around a constant value, it is likely that the stationary phase has been reached. This process should be repeated for each of the estimated parameters in the analysis to ensure that apparent convergence of likelihood reflects convergence of all parameters before the run can be assumed to have converged.

2.5.3 Effective sample size

As each iteration of a Markov chain proposes a new value for a single parameter, successive iterations are not independent. Sampling every 100 or 1000 iterations reduces the correlation between iterations in the sample of the MCMC and the degree of independence can be quantified as the effective sample size (ESS). The ESS of an analysis is the number of effectively independent draws from the Markov chain that the sample is equivalent to. An ESS of >100 (ideally >200; Drummond & Rambaut 2004) should be sought to ensure that the sample from the stationary phase is representative of the target distribution. If samples are taken too frequently, they will be correlated and the ESS will not be improved. The most reliable ways to increase ESS are to increase run length and to combine the results of two or more independent runs.

2.5.4 Potential scale reduction factor

The potential scale reduction factor (PSRF; Gelman & Rubin 1992) is an estimate of the factor by which the scale of a parameter may be reduced under an infinite number of iterations. As PSRF approaches 1, the parameter estimates can be assumed to have converged.

2.5.5 Parameter density plots

Histograms of the frequency of values in the sample of the MCMC show the distribution of values. The ideal sample will be normally distributed with low variance, although in practice a unimodal distribution is usually acceptable. Estimates of the 95% upper and lower confidence intervals on the mean of a parameter indicate the range of possible 'real' values for that parameter.

2.5.6 Convergence diagnostics for this thesis

Since MCMC convergence is essential to the validity of BI, a combination of the above convergence diagnostics will be used in this thesis to ensure that each MCMC run has converged. If the majority of diagnostics do not show convergence, the problem will be resolved using one or more of the following steps: (i) increase run length; (ii) remove problematic model parameters if the model appears over-specified and a simpler model receives sufficient support (e.g. is within the 95% set of models using AIC weights) during model selection (Section 2.3.3); and (iii) adjust the temperature parameter of the Metropolis coupling to increase the rate at which chains exchange parameters.

2.6 Node support indices

A number of indices have been proposed to measure the support for individual nodes within a phylogenetic topology. Two of the most widely used, bootstrap and posterior probability, are discussed below.

2.6.1 Bootstrap

In selecting a single topology, MP, NJ and ML do not provide a measure of confidence in the resulting phylogeny. Several methods have been proposed to address this issue, of which the most commonly applied is the non-parametric bootstrap (Efron 1979), first applied to phylogenetics by Felsenstein (1985). The data used to generate the original phylogeny are resampled with replacement to generate at least 100 pseudoreplicate datasets of the same length as the original. The

same phylogenetic method is applied to these datasets and the proportion of times each clade is found is recorded and used to gauge the level of support for the recovered clades in the original phylogeny. The value of the non-parametric bootstrap in assessing the level of support within the data for a given phylogenetic hypothesis lies in the random resampling regime. Each pseudoreplicate dataset contains the same total number of characters as the original. However, individual characters may be repeated or absent, so any relationship that is only supported by a small number of characters will only be supported by those pseudoreplicate datasets that contain those characters. With a single ML analysis taking many hours or even days (depending on the size of the dataset and speed of computer used), the time required to apply this technique can be prohibitive. An important condition of the bootstrap is that while it may take a large number of characters to obtain high confidence on short internal branches (Berbee *et al.* 2000), this high confidence may be an artefact of the failure of the resampling regime to randomise large datasets.

2.6.2 Posterior probability

BI through MCMC sampling provides an estimate of confidence (posterior probability) in the course of phylogenetic reconstruction, which is more computationally efficient than running bootstrap replicates. The principal advantage of Bayesian posterior probabilities is particularly evident in short sequence alignments where each character is assessed independently exactly once. While short internal branches receive weak support from bootstrap replicates of short sequence alignments, due to the resampling regime 'ignoring' the small number of variable characters in many of its replicates, MCMC samples are able to assign high posterior probability values to these clades (Alfaro *et al.* 2003). However, the high value of posterior probabilities relative to bootstrap values has been criticised and posterior probabilities should be interpreted conservatively.

2.6.3 Support indices for this thesis

The selection of support indices cannot be separated from the selection between Bayesian and non-Bayesian phylogenetic reconstruction techniques. The choice of BI for tree reconstruction in this thesis (Section 2.2.5) is based, in part, upon the ability to calculate node support from the entire sequence alignment, which is not possible with the bootstrap.

2.7 Testing the validity of a tree framework

2.7.1 Problems with phylogenetic trees

The modern concept of an evolutionary tree was introduced by Darwin (1859) and remains the most widely recognised and applied representation of the evolutionary relationships among a set of taxa. A phylogenetic tree can be considered to be a special class of network in which each taxon partition is represented by a single edge (a 'branch' in conventional tree terminology). Two features of sequence evolution present problems to the representation of phylogenetic relationships on trees (Huson & Bryant 2006): (i) homoplasy, leading to uncertainty in the order of accumulation of mutations; and (ii) reticulate events, such as recombination, horizontal gene transfer or hybridisation, leading to multiple evolutionary trees within a single sequence alignment.

2.7.2 Building networks

Phylogenetic networks provide methods to analyse and visualise non-treelike data. Ideally the choice of method should be governed by the phylogenetic history of the taxa under investigation. Phylogenetic uncertainty due to homoplasy can be represented implicitly using split decomposition in which each split (a partition of all taxa into two subsets) is represented by one or more parallel edges. Early implementations of this method (Bandelt & Dress 1992) have been superseded by model based methods (Bryant & Moulton 2004) with extra topology-related parameters to reduce systematic error. Methods have also been developed to produce reticulate networks which provide an explicit representation of the evolution of

recombining sequences within a set of taxa. To date such methods have only been applied to recombination of binary data (Gusfield & Bansal 2005; Huson & Klöpper 2005).

Alternate methods that make fewer assumptions regarding the underlying cause of the departure from tree-likeness, such as statistical parsimony networks (Clement *et al.* 2000), continue to be widely used in phylogeography as they can be used in nested clade phylogeographic analysis (NCPA; Templeton 1998). NCPA provides a statistical framework for interpreting phylogeographic distributions, however, since relatively arbitrary criteria are typically used to break loops (Crandall & Templeton 1993; Pfenninger & Posada 2002) it can be considered to implicitly utilise a tree framework.

2.7.3 Looking for trees in networks

The hypothesis that the sequences originated on a phylogenetic tree is nested within the hypothesis that the sequences originated on a network so it is possible to conceive a statistical framework to test for tree-like evolution. One such test is based on the principle illustrated in Figure 2.4. Starting with a split network N inferred from a dataset, the test to determine whether it is likely that the data originated on a tree involves two steps (Huson & Bryant 2006): (i) construct a 95% confidence network for N to obtain the 95% confidence limits of the length of each edge; and (ii) if the confidence network does not contain a tree (see Figure 2.4) then reject the null hypothesis that the data originated on a tree. Simulations have been used to demonstrate that this test is valid, but has unacceptably low statistical power (Huson & Bryant 2006).

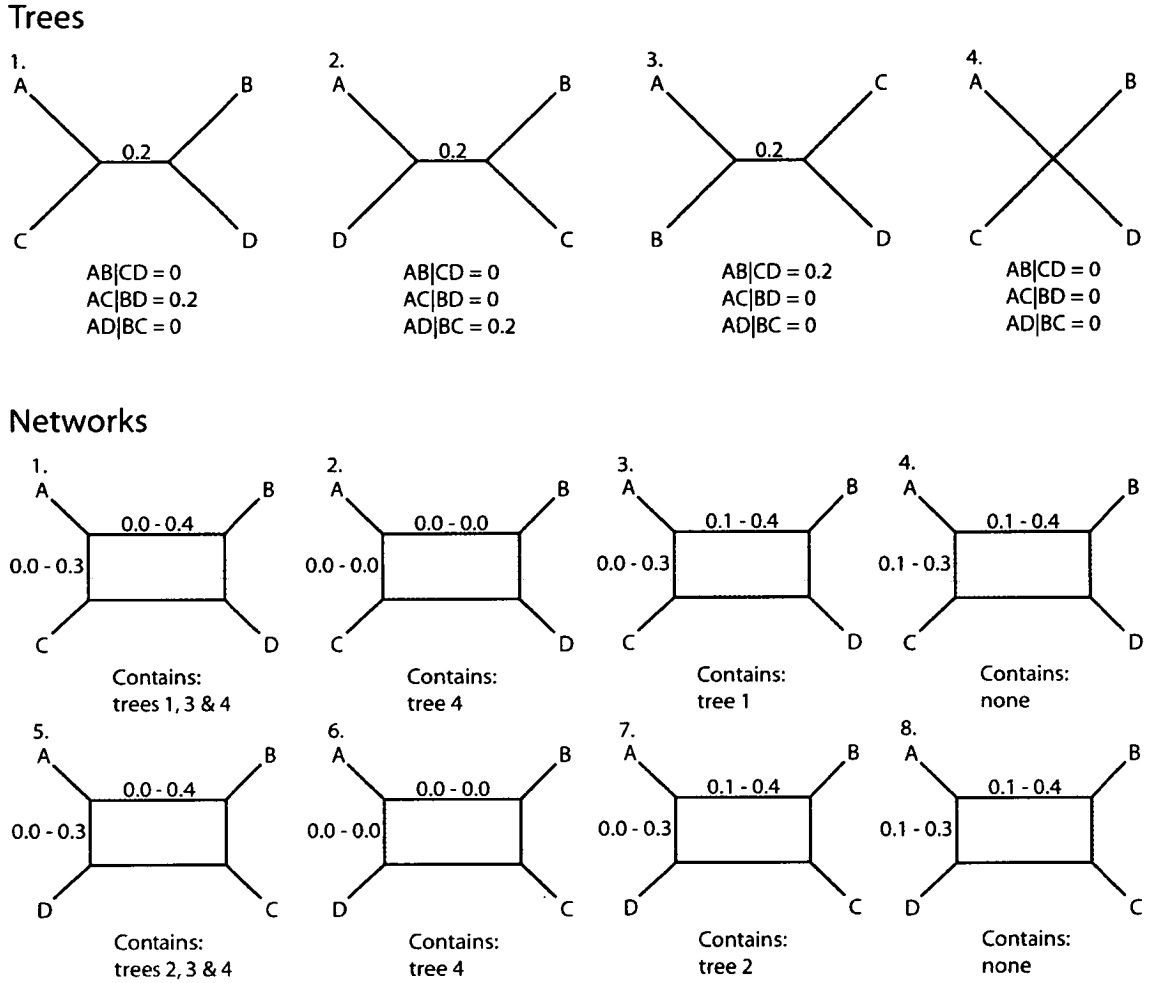


Figure 2.4 For a set of four taxa a total of four unrooted (three resolved and one unresolved) tree topologies are possible. In a set of confidence networks for the same four taxa, networks 1-3 and 5-7 each contain at least one tree, although none of these networks supports all four possible tree topologies. In contrast, networks 4 and 8 do not contain any tree since no tree topology is possible for four taxa in which two splits have a value greater than zero.

2.7.4 Phylogenetic framework for this thesis

The prospect of a statistical framework to allow an objective choice between tree and network frameworks is encouraging, however, the low statistical power of the test at present means that the choice between frameworks is likely to remain controversial. This thesis will make use of both phylogenetic networks and trees. Trees will be used to reconstruct interspecific relationships, while the choice of frameworks for intraspecific phylogenies will be guided in part by the test of Huson & Bryant (2006). In recognition of the limitations of this test, and because departures from

tree-likeness in mitochondrial DNA are unlikely to be due to true reticulate events, a tree framework will be adopted for reconstruction of some intraspecific phylogenies in order to make use of techniques (Sections 2.8-2.9) that cannot be applied to networks.

2.8 Molecular clock

The strict molecular clock hypothesis assumes that the rate of molecular evolution is constant through time and across branches. If this assumption can be met, molecular clocks can provide powerful evidence for dates and relative depths of nodes within phylogenies.

2.8.1 Testing the assumption

The two extremes of the molecular clock hypothesis are: (i) that the rate of sequence evolution is constant throughout a tree (strict clock) and (ii) that every branch has an independent rate of molecular evolution (no clock). Each of these hypotheses is relatively simple to implement compared with any intermediate hypothesis and, as such, phylogenies have traditionally been generated either under a strict clock or in the absence of a clock. Tests of the validity of a molecular clock are easy to implement using either LRTs (Section 2.3.3.3) or Bayes factors (Section 2.3.3.4), depending on the framework of the investigation. In such tests, the molecular clock hypothesis (all branches have the same rate of evolution) is a special case of the non-clock hypothesis (all branches have an independent rate of evolution) so the hypotheses are nested. The use of a less appropriate model in phylogenetic reconstruction is expected to result in an increase in the ln$L$ or marginal ln$L$ of the resulting phylogenetic hypothesis.

2.8.2 Relaxed clocks and relaxed phylogenetics

Alternatives to the strict clock have included assigning branches to user defined sets with different rates of sequence evolution, but such methods are only useful given a strong *a priori* reason to believe specific groups will differ (Yoder & Yang 2000).

An alternative method, Bayesian relaxed-clock models (Thorne *et al.* 1998), remove the need for user defined rate sets but can only be applied to predefined topologies. This is a problem since the assumption of a molecular clock (whether strict or relaxed) is likely to affect the topological estimate, so optimal relaxed clock phylogenies cannot be determined with this method (Drummond *et al.* 2006). The new field of "relaxed phylogenetics" (Drummond *et al.* 2006), implemented in BEAST (Section 2.12.3), offers a framework in which phylogeny and divergence times can be co-estimated under relaxed-clock models.

2.8.3 MRCA versus MACA

The degree of certainty of the age of a clade is typically estimated by consideration of the 95% confidence intervals for the age of the **most recent common ancestor, MRCA**, of the clade. Hayward & Stone (2006) point out that this most recent possible date is only one extreme of the range of possible dates. Depending on demographic processes that have occurred since divergence of the clade from its nearest sister clade, the actual range of ages for the clade lies between the MRCA date and the date range of the **most ancient common ancestor, MACA** (Figure 2.5).

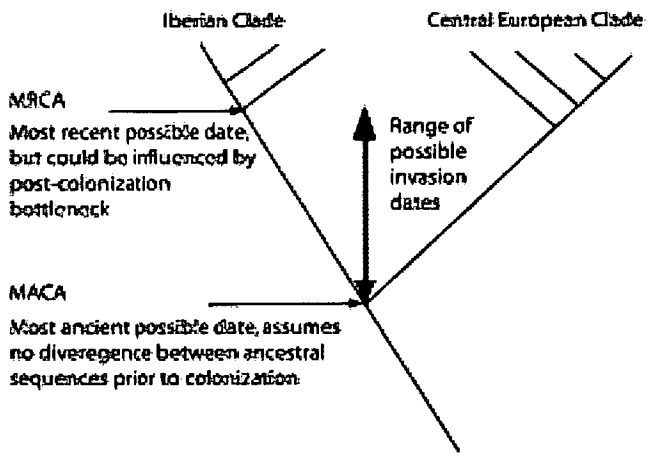


Figure 2.5 The assumptions of MRCA and MACA estimation (taken from Hayward & Stone 2006).

2.8.4 Molecular clocks for this thesis

Ideally, molecular clocks should be calibrated with data such as fossil evidence for the taxa under investigation. No such calibration data are available for the

Cynipinae, so divergence will be calibrated using the widely applied approximation for mitochondrial DNA of 2.3% sequence divergence per million years (Brower 1994). Use of such an approximate calibration means that date estimates should be interpreted with caution, however, the accuracy of relative time depth of nodes is not affected. Since BI is the preferred phylogenetic reconstruction technique (Section 2.2.5), the validity of the molecular clock hypothesis will be tested using BF (Section 2.3.3.4). For all nodes for which the date is estimated, both MRCA and MACA ages will be calculated.

2.9 Trait reconstruction

2.9.1 Input trees

Traditional trait reconstruction methods, using MP (Section 2.2.1) or ML (Section 2.2.3), map trait evolution onto a single input tree. The only way to account for topological uncertainty is to multiply the probability of a given character state at a node by the bootstrap proportion or posterior probability support for the node. This approach does not account for the fact that reconstruction at well supported nodes is likely to be affected by changes in the topology elsewhere in the phylogeny. Early attempts to address this issue involved mapping traits onto sets of trees obtained through bootstrapping (Section 2.6.1) of trees or equally parsimonious trees (e.g. Hibbett *et al.* 2000). These approaches are thought to reflect uncertainty, but the statistical properties of the tree sets are unclear (Felsenstein & Kishino 1993).

2.9.2 Trait scoring

For discrete characters, trait scoring is straightforward. However, there are many situations, particularly in areas such as phylogeographic reconstruction, in which determining an appropriate scoring regime can be difficult. In such situations, a balance must be sought between resolution and potential over-parameterisation. Adding trait categories will only increase resolution if the resulting set of trait categories is genuinely distinct.

2.9.3 Trait mapping for this thesis

This thesis will use trait mapping on sets of trees sampled during MCMC simulation, since these tree sets should reflect topological uncertainty in a statistically meaningful way. Details of the methods used to map phylogeographic distributions will be discussed in Chapter 3.

2.10 Molecular operational taxonomic units (MOTUs)

2.10.1 Reasons for choosing MOTUs

The terminal nodes of phylogenetic trees and networks are typically labelled with taxon names. However, this raises two important issues: (i) if the taxon labels are used to determine character-states for subsequent trait mapping, the accuracy of the trait mapping will be limited by the accuracy of taxon delimitation; and (ii) while incomplete taxon sampling should be avoided where possible, if a group of individuals are represented in a phylogenetic analysis by a single sequence per gene, it is essential to ensure that it is biologically meaningful to consider them to be members of a single group. The concept of a molecular operational taxonomic unit (MOTU; Floyd *et al.* 2002) has been developed in recognition of the limitations of morphological taxonomy, both in terms of the limited numbers of trained taxonomists (Lawton *et al.* 1998) and the frequency of cryptic variation between taxa (Blaxter 2004). The aim of MOTU definition is to assign organisms to approximately species- or genus-level groups on the basis of molecular sequence data (Floyd *et al.* 2002). MOTUs are commonly used to assign samples to groups on the basis of molecular barcode data, for example cytochrome *c* oxidase subunit I (*coxI*) sequences. While they are frequently applied to culture-independent studies of prokaryotes (Pace 1997), MOTUs can also be valuable in studies of morphologically identified taxa, where they may reveal evidence of cryptic speciation.

2.10.2 Defining MOTUs

Commonly applied objective criteria for defining MOTUs fit into three categories: (i) phylogeny based: (ii) BLAST match based; and (iii) cluster based. Phylogeny based definitions are closely allied to phylogenetic species concepts – sequences are assigned to the same MOTU if they are members of the same monophyletic clade. The requirement for a phylogeny means that this approach is computationally intensive and, for datasets that cannot be considered tree-like, monophyly may be difficult to define on a network. The use of BLAST algorithms to determine taxon identity by the identity of the highest scoring match in the NCBI database is often restricted by a lack of sufficiently closely related sequences (Blaxter *et al.* 2005). Cluster based methods allow MOTU designation based on a set of sequences without any requirement for previously identified specimens. A local BLAST database can be created by sequentially adding each of the sequences and assigning those with less than a predefined cut-off number of base differences to MOTUs (Blaxter *et al.* 2005). Clustering methods are influenced by the order of sequence addition so it is important to randomly resample the sequences a large number of times to determine which sequences are consistently placed into the same MOTUs (Blaxter *et al.* 2005).

2.10.3 MOTUs for this thesis

MOTUs for this thesis will be defined using clustering methods. The disadvantages to clustering will be reduced by resampling and defining MOTUs at a range of cut-offs.

Part 2 – Molecular and bioinformatic methods

2.11 Laboratory protocols

2.11.1 Samples

Details of sample collection locations are presented in each chapter. Oak gallwasps and oak inquilines were reared from their galls under quarantine in Edinburgh.

Samples for DNA sequencing were preserved from life in ethanol and were stored at -20°C. Samples for allozyme screening were stored at -80°C.

2.11.2 DNA extraction

DNA was extracted from whole wasps using the DNeasy Tissue Kit (QIAGEN cat. 69504), following the manufacturer's protocol for insect DNA extraction, or from individual legs using Chelex-based DNA extraction.

2.11.3 PCR amplification

For all genes, 25µl polymerase chain reactions (PCRs) were carried out in a PTC-200 DNA Engine (MJ Research) using 1 U *Taq* polymerase (Invitrogen or Promega), 2.5 µl 10x *Taq* buffer, 1.5 µl $MgCl_2$ (25 mM), 0.5 µl dNTPs (10 mM), 0.35 µl primers (20 pmol), 1.0 µl template DNA and 18.85 µl dH_2O. Specific details of primer sequences are presented in Table 2.1.

Table 2.1 Primer sequences for each of the four genes used in this thesis. Cytochrome *b* was typically amplified using the CB1/CB2 primer pair. If this failed, an overlapping fragment was amplified using the CP1/CP2 primer pair.

Gene	Abbrev.	Primer (F/R)	Sequence (5'-3')	Reference
Cytochrome *b*	*cytb*	F, CB1 R, CB2	TATGTACTACCATGAGGACAAATATC ATTACACCTCCTAATTTATTAGGAAT	Jermiin & Crozier (1994)
Cytochrome *b*	*cytb*	F, CP1 R, CP2	GATGATGAAATTGGATC CTAATGCAATAACTCCTCC	Harry *et al.* (1998)
Cytochrome *c* oxidase subunit I	*coxI*	F, HCO2198* R, LCO1490	TAAACTTCAGGGTGACCAAAAAAT GGTCAACAAATCATAAAGATATTGG	Folmer *et al.* (1994)
Nuclear 28S ribosomal gene, D2 region	28SD2	F R	CGTGTTGCTTGATAGTGCAGC TCAAGACGGGTCCTGAAAGT	Hancock *et al.* (1988)
Nuclear 28S ribosomal gene, D3-5 region	28SD3-5	F R	ACACACTCCTTAGCGGA GACCCGTCTTGAAACACGGA	Friedrich & Tautz (1995)

2.11.4 DNA sequencing

PCR products were purified using shrimp alkaline phosphatase and *E. coli* exonuclease I (USB Corporation, USA) and sequenced directly on an automated ABI Prism 3730 Genetic Analyzer machine using ABI BigDye v3.1 Terminator Sequencing chemistry. All PCR products were sequenced in both directions to

minimise PCR artefacts, ambiguities and base-calling errors. Chromatogram output was checked by eye using Sequencher 4.1 (Gene Codes) or ProSeq (Filatov 2002).

2.11.5 Cloning

Direct sequencing of a small proportion of *cytb* PCR products revealed multiple *cytb*-like fragments, or sequences possessing reading frames containing stop codons or indels, suggesting the possible presence of nuclear pseudogenes (Bensasson *et al.* 2001). In these cases, individual PCR products were amplified by cloning (TA cloning, Invitrogen) and only specimens for which a single correct open reading frame (ORF) bearing sequence was identified have been included in the following analyses.

2.12 Software

The assumptions and methodological details of the software used in this thesis are discussed below for each program in turn.

2.12.1 Arlequin 3.0 (Schneider *et al.* 2000)

Arlequin provides tools for population genetic analysis. In this thesis it will be used to: (i) test for Hardy-Weinberg equilibrium; (ii) produce pairwise mismatch distributions; and (iii) calculate summary statistics for DNA sequence data.

2.12.1.1 Hardy-Weinberg equilibrium

Under the assumption that a population of genes is infinitely large and randomly mating with no selection or mutation, the Hardy-Weinberg equilibrium predicts that allele frequencies will remain constant. For a gene with two alleles, A and B, with frequencies p and q, respectively, the frequencies of each of the three possible genotypes (AA, AB and BB) in the next generation are given by the equation:

$$(p + q)^2 = p^2 + 2pq + q^2$$

Departures from the expected frequencies indicate departures from the assumptions listed above and can be tested using Fishers exact test on a 2x2 contingency table. Arlequin tests for significant departures from Hardy-Weinberg equilibrium using the approach of Guo & Thompson (1992), which can accommodate genes with more than two alleles.

2.12.1.2 Pairwise mismatch distributions

A mismatch distribution is produced by plotting the relative frequency of each number of nucleotide differences between all possible pairs in a set of sequences. Under the constant population size model the expected mismatch distribution is a multimodal ('ragged') exponential decline in frequency. Under the population growth-decline model the distribution is expected to be a unimodal ('bell-shaped') curve (Slatkin & Hudson 1991). Comparison of data under these models allows inference of population demographic histories (Emerson *et al.* 2001). Mismatch distributions will be tested for unimodality using the sum of squared deviations (SDD) test of Schneider & Excoffier (1999). For lineages showing significantly unimodal distributions, relative population sizes ($2\mu N$) before (θ_0) and after (θ_1) population growth, and the relative time since the onset of population expansion ($\tau=2\mu t$) can be estimated.

2.12.1.3 Summary statistics

Summary statistics describe basic properties of a sequence set, such as nucleotide diversity, π (the sum of the proportion of different nucleotides between each pair of sequences), and haplotype diversity, h (which depends on the number of haplotypes and the relative frequency of each haplotype).

2.12.2 BayesMultiState (Pagel *et al.* 2004)

BayesMultiState provides methods to account for topological uncertainty in trait reconstruction by mapping character-state evolution onto a set of trees produced by MCMC sampling (Section 2.9.1). Both ML and BI trait mapping are available but

ML is preferred for this thesis due to the difficulty of defining objective priors (Section 2.4) for trait mapping.

2.12.3 BEAST 1.4 (Drummond & Rambaut 2006)

BEAST provides methods for simultaneous BI of rooted phylogenies and estimation of node ages. This thesis uses strict clock methods with a coalescent prior on branch lengths to estimate ages of MRCAs and MACAs (Section 2.8.3) of specific clades. Input files for BEAST will be generated using the accompanying program, BEAUti.

2.12.4 ClustalX (Thompson *et al.* 1997)

ClustalX provides methods to align sequences. ClustalX uses a relatively basic sequence alignment algorithm and will only be used for alignment of sequences of similar length. Sequences are aligned to a guide tree generated from a distance matrix of the unaligned sequences. In order to align sequences of different lengths, gaps must be introduced to the alignment. Penalties for opening and extending gaps to produce the alignment can be varied at different places in the sequence. Datasets with greater sequence alignment uncertainty will be aligned using MUSCLE (Section 2.12.10).

2.12.5 DnaSP 4.10 (Rozas *et al.* 2003)

DnaSP provides tools for population genetic analysis. In this thesis it will be used to: (i) calculate Fu's *Fs* statistic; (ii) perform the McDonald-Kreitman test for selection; and (iii) produce pairwise mismatch distributions.

2.12.5.1 Fu's F_S statistic

Fu's F_S statistic is used to detect population growth. In an expanding population, it has a large negative value due to the excess of singleton mutations compared with the expectation for a static population (Fu 1997).

2.12.5.2 McDonald-Kreitman test

The McDonald-Kreitman test uses the ratio of non-synonymous (affect amino acid coded for) to synonymous (do not affect the amino acid coded for) mutations, the d_N/d_S ratio. In a selectively neutral gene, this ratio is expected to be one. Higher values indicate diversifying selection, while lower values indicate purifying selection.

2.12.5.3 Pairwise mismatch distributions

Within DnaSP, the departure of a mismatch distribution from unimodality is calculated using the raggedness statistic, *r* (Harpending 1994). However this has low statistical power and additional test statistics such as SSD (Section 2.12.1.2) or Fu's F_S (Section 2.12.5.1) should also be calculated.

2.12.6 Genetix 4.0 (Belkhir 1999)

Genetix provides tools for population genetic analysis. In this thesis it will be used to test for linkage disequilibrium.

2.12.6.1 Linkage disequilibrium

Alleles for two unlinked genes in a population would be expected to occur independently. Linkage disequilibrium, *D*, describes the increased frequency with which pairs of alleles for the two genes are observed, due to physical (e.g. both genes are on the same chromosome) or functional linkage.

2.12.7 GEODIS 2.4 (Posada *et al.* 2000)

GEODIS performs statistical analysis of the association between haplotype nesting and geographic distribution for NCPA. The results of this are interpreted using the most recent version of the key of Templeton *et al.* (1995; available at http://inbio.byu.edu/Faculty/kac/crandall_lab/geodis.htm).

2.12.8 ModelTest 3.6 (Posada & Crandall 1998)

ModelTest performs hLRT, AIC, and BIC (Section 2.3.3) tests on a sequence alignment. The procedure has two stages: (i) a PAUP* (Section 2.12.11) command block is used to estimate the likelihood of the data under each of 56 models given a phylogenetic tree (which can be user defined or estimated in PAUP*); (ii) the model likelihoods are read into ModelTest to determine the best model(s). In this thesis, a modified version of ModelTest, MrModelTest (Nylander 2004), which only compares models that can be defined in MrBayes (Section 2.12.9) will be used to select models for BI.

2.12.9 MOTU_define 2.04 (Floyd & Blaxter 2006)

MOTU_define clusters input sequences into MOTUs (Section 2.10) by adding each sequence in turn to a local BLAST database and then taking the next sequence and performing a Blast similarity search against the entries in the local database. If the sequence has less than a user-defined cut-off number of differences to an existing MOTU cluster then it is added to that cluster, otherwise it is assigned to a new cluster. The grouping of sequences in this way is sensitive to the order in which they are added (Blaxter *et al.* 2005) so 100 replicates using different random resampling orders will be performed for each MOTU cut-off. The MOTU_define perl script was modified to fix a bug in the randomisation routine and to allow the MOTU_define to run on Windows. These changes have been incorporated into MOTU_define 2.07.

2.12.10 MrBayes 3.1 (Ronquist & Huelsenbeck 2003)

MrBayes produces both rooted and unrooted BI (Section 2.2.4) phylogenies and separate models can be applied to each of two or more data partitions. Details of specific MC^3 parameter settings are provided in the relevant chapters.

2.12.11 MUSCLE (Edgar 2004)

MUSCLE aligns sequences in a three stage process: (i) draft progressive alignment, a UPGMA guide tree is inferred from a *k*mer (contiguous subsequences of length *k*)

distance matrix generated from the unaligned sequences and sequences are aligned using the guide tree; (ii) improved progressive alignment, step one is repeated using the aligned sequences as an input and the distance matrix is calculated using Kimura distances, which are corrected for multiple substitutions at a single site; and (iii) refinement. For the first cycle of stage three, a guide tree based on the step two alignment is split into two subtrees. These are used to produced two separate partial sequence alignments which are then combined to give a full sequence alignment. For subsequent cycles of step three, the sequence alignment from the previous cycle is used in place of the step two sequence alignment. If the sequence alignment has a lower SP (sum of pairwise alignments) score it is retained, otherwise the previous sequence alignment is retained. Step three is continued until convergence is reached.

2.12.12 PAUP* 4.10b (Swofford 2001)

PAUP* provides methods for phylogenetic analysis by MP (Section 2.2.1), NJ (Section 2.2.2) and ML (Section 2.2.3). PAUP* will not be used for phylogenetic reconstruction but be used to estimate likelihoods of datasets under different models and tree topologies.

2.12.13 SplitsTree 4.4 (Huson & Bryant 2006)

SplitsTree will be used to produce phylogenetic networks. Networks will be constructed using the NeighborNet (Bryant & Moulton 2004) distances transformation on model-based distances and equal angle splits transformation (Dress & Huson 2004). The 95% confidence networks for the test of tree-likeness (Section 2.7.4) will be calculated from 1000 bootstrap replicates.

2.12.14 Structure (Pritchard *et al.* 2000)

Structure provides methods to assign individuals to populations on the basis of genotype data and will be used in the analysis of Allozyme data. Structure assumes a model in which a specified number of populations are characterised by a set of allele frequencies derived from multilocus genotype data. Individuals are assigned

probabilistically to populations under the assumptions of Hardy-Weinberg (Section 2.12.1.1) and linkage equilibrium (Section 2.12.6.1) using MCMC simulation. The simulation is rerun for models specifying different numbers of populations (K) and the posterior probabilities for each simulation are compared to infer the number of discrete populations best supported by the data. Structure allows fitting of models with or without admixture, the latter allowing genotypes to arise through mating between individuals derived from different populations.

2.12.15 TCS 1.2.1 (Clement *et al.* 2000)

TCS generates statistical parsimony haplotype networks. Statistical parsimony networks are used to perform NCPA.

2.12.16 TreePuzzle (Schmidt *et al.* 2002)

TreePuzzle provides methods to perform likelihood mapping. Likelihood mapping (Strimmer & von Haeseler 1997) provides an estimate of the phylogenetic utility of a set of sequences. For each group of four sequences (quartets) within the sequence set, there are three possible topologies. The likelihood of each of these hypotheses is plotted on three axes arranged at 120° to each other from the centre of an equilateral triangle. Points plotted in the regions nearest to the corners represent fully resolved quartets, points along the edges of the triangle represent quartets that provide equal support for two of the three topologies, and points in the central region represent unresolved, star-like quartets. The distribution of points once all possible quartets have been plotted reflects phylogenetic utility.

Chapter 3

Longitudinal phylogeography of the Western Palaearctic: patterns and methodology

Chapter 1 introduced the previous work on oak gallwasp phylogeography. One of the emerging patterns was the suggestion of eastern origin and common longitudinal gradients in diversity. This chapter reviews the phylogeographic literature to determine the extent to which the patterns noted in the oak gallwasps can be applied to more diverse Western Palaearctic taxa.

3.1 Introduction

Phylogeographic studies within the Western Palaearctic have typically concentrated on latitudinal gradients in diversity that have resulted from post-glacial colonisation of northern Europe from southern refugia (Section 1.2.3). Such studies frequently involve taxa with distributions that extend across the entire Western Palaearctic and beyond, however sampling regimes in many studies appear biased towards European populations. In considering only the most recent post-glacial dispersal of a species, latitudinal phylogeography treats refugia as sources of diversity rather than as sets of related populations. This chapter attempts to establish the extent to which the same data can be used to address questions from a longitudinal perspective, across multiple glacial cycles.

The first aim of this chapter is to present a literature survey to establish the origins of widespread Western Palaearctic species, with particular emphasis on the role of little-sampled regions to the east of Europe. For taxa sampled across their full longitudinal range, putative origins identified either on the basis of phylogenetic reconstruction or patterns in the diversity of unrootable markers are surveyed. A set of criteria (see Section 3.2) are applied to each analysis assessing the strength of the support for a given geographic origin. These are deliberately conservative, so that our conclusions are based on studies showing well-supported patterns. Specifically,

this chapter asks whether available data more commonly support one of the following hypotheses, each of which could operate at a range of timescales:

(i) *Out of the east.* That widespread species originated to the east of Europe, and then dispersed westwards. If supported, we also ask whether extra-European populations are genetically distinct from those in Europe. Sampling restricted to Europe in such cases would miss significant within-species genetic diversity, with implications for assessment of conservation priorities.

(ii) *Central European diversification.* That widespread species originated in central Europe (Italy, Hungary, the Balkans) and then dispersed both eastwards and westwards.

(iii) *Out of the west.* That widespread species originated in western Europe and dispersed eastwards beyond Europe. If supported, we also ask whether extra-European populations are genetically distinct from those in Europe. Sampling restricted to Europe will only provide accurate assessment of conservation priorities if extra-European populations are both derived relative to Europe, and only contain a subset of the genetic diversity present in Europe.

Each of these hypotheses could operate at a range of timescales (from Holocene through Pleistocene to pre-Pleistocene; Section 1.2.2). While latitudinal phylogeography considers populations that contain a subset of the genetic diversity of a source population, the range of timescales over which longitudinal phylogeographic processes can act raises the possibility that populations will have time to accumulate regionally specific genetic polymorphisms. Longitudinal processes of range expansion during interglacials and isolation during glacial maxima have the potential to generate levels of genetic differentiation of direct relevance to conservation, including differentiation into evolutionary significant units, sub-species, sibling species and even radiations within genera. Inferring the timescale of origin and dispersal may be possible at much finer resolution,

identifying specific interglacial periods during which phylogeographic events occurred. However, the resolution that can be achieved in individual studies will depend upon the evidence available to calibrate timescales. Some problems with dating are discussed in Section 3.5.5. A further challenge in the reconstruction of ancient phylogeographic processes is the fact that multiple longitudinal range expansion processes may have contributed to the present phylogeographic distribution, with subsequent refugial persistence or extinction.

The second aim of this chapter is to provide an overview of methodology appropriate to distinguishing among alternative longitudinal dispersal hypotheses (Section 3.5).

3.2 Materials and methods

This section describes the criteria for selection of papers for consideration in this review (Section 3.2.1) and provides further criteria for robust interpretation of the results of the studies in light of some potential artefacts in longitudinal phylogeographic reconstruction (Section 3.2.2) and the timescale of each study (Section 3.2.3).

3.2.1 Literature review

Published research articles were selected for the survey by searching the Web of Science (http://wos.mimas.ac.uk/) using the search terms: 'eur* AND phylogeo*'; 'palaearctic AND phylogeo*'; and 'geographic origin AND palaearctic'. The results were screened for studies that: (i) used phylogeographic methods to investigate patterns in taxa whose distributions extend from Europe eastwards across the Western Palaearctic (including studies that also extended to northwest Africa); (ii) sampled a large portion of the geographic range, since studies that focus on a highly localised area usually have specific aims that are not dependent upon coverage of the entire range; and (iii) sampled taxa at approximately the species level, since longitudinal phylogeography requires at least species level coverage but studies

above the sibling species level typically have insuffient geographic sampling for each individual species. Each of the studies that met the criteria of this review was examined to identify: (i) the known distribution of the taxon involved; (ii) the extent of the range that had been sampled; (iii) the types of molecular markers used (Section 2.1); (iv) the timescale of the most ancient and most recent common ancestors (MRCA and MACA; Section 2.8.3); and, for studies in which a geographic origin was suggested, (v) the area of origin. Where the geographic origin was not reported in the study, studies were assigned to the following categories: (i) incomplete sample coverage; (ii) lack of appropriate rooting; and (iii) lack of phylogeographic structure in the marker used.

3.2.2 Interpreting longitudinal patterns

For those studies in which an area of origin was inferred, a robust set of criteria were developed to assess the strength of evidence from a longitudinal perspective. Longitudinal patterns are most likely to be resolvable in widespread taxa whose populations are old relative to the age of the taxon. This reduces the chance that the current phylogeographic distribution has been affected by localised extinction of basal clades. The assumption that current distributions of diversity reflect initial colonisation patterns is most likely to be valid for datasets in which the MACA (Section 2.8.3) of each sampled population is of a comparable age to the MACA of the taxon (Figure 3.1a-c). It is important to consider the MACA rather than the MRCA to avoid incorrect rejection or acceptance of a dataset. A recent MRCA for a population where the MRCA of the taxon is relatively ancient could reflect a population bottleneck and would not preclude the use of the dataset for longitudinal phylogeography provided the MACA for the population is sufficiently old (Figure 3.1b). Conversely, a dataset may appear acceptable if the MRCAs of the populations are of a comparable age to the MRCA of the taxon but if the MACA of the taxon is relatively ancient the distributions may have been affected by local extinctions (Figure 3.1d). Local extinctions in individual regions could also leave a phylogeographic signature that could not be distinguished from a scenario in which some regions had only recently been colonised for the first time (Figure 3.1e-f).

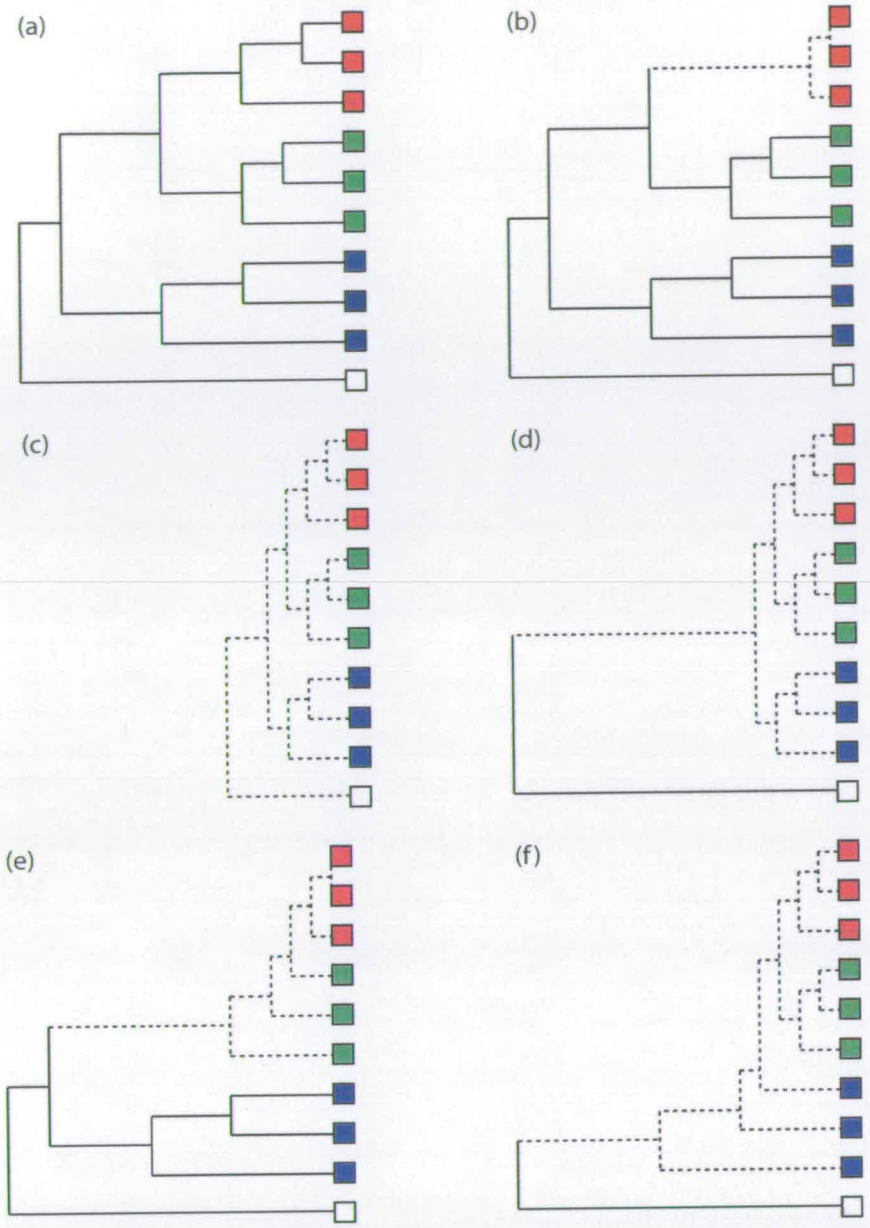


Figure 3.1 The effect of species' history on phylogeographic reconstruction. Dashed lines indicate branches that are changed relative to scenario a. An outgroup sequence is represented by a white square and samples from three regions are indicated by red, green and blue squares. Three scenarios allow the aims of longitudinal phylogeography to be addressed: (a) ancient species origin and ancient range expansion; (b) population bottleneck, reduces diversity in the affected region but does not affect the reconstructed relationships between regions, which are as in (a); (c) recent species origin and recent range expansion, which is supported by the closely related outgroup taxon. Three scenarios complicate accurate inference of longitudinal phylogeography: (d) recent origin and range expansion with a distantly related outgroup or widespread extinction and recent recolonisation. The geographic locations of unsampled or extinct lineages of the target taxon are expected to have a significant impact on inferred phylogeographic patterns; (e) local extinction following colonisation affects relationships between regions and cannot be distinguished from a scenario in which the affected region had not been previously colonised; and (f) as for (e) but affecting multiple regions.

3.2.3 Timescales of range expansion

Holocene (Section 1.2.2) range expansion predicts a single centre of genetic diversity, with other parts of the longitudinal range possessing a subset of this diversity and with little or no refuge-specific polymorphism.

Late Pleistocene range expansions predict a centre of diversity from which others are derived over timescales that allow the evolution of refuge-specific polymorphism. In this case, descendant populations in southern refugia are expected to have haplotypes that are phylogenetically derived from more basal sequences associated with their origin, but which may be refuge specific. This pattern will overlay sorting of ancestral polymorphism, whose persistence will depend on the mutation rate of the marker, and on the size of longitudinally dispersing populations.

Early or pre-Pleistocene range expansion events are likely to have resulted in ancient founding of populations in southern glacial refugia, which, because of their age, are likely to have completed sorting of ancestral polymorphism.

The requirement for corroborating evidence in support of a species origin depends upon the longitudinal pattern (Section 3.2.2) and the timescale of range expansion. Support that an inferred Holocene expansion is the single major longitudinal range expansion event for a taxon (rather than the most recent in a series) comes from evidence such as the presence of a sister/sibling species in the centre of diversity (effectively removing the possibility of artefacts associated with lineage extinctions). The inference of pre-Holocene eastern origin can be considered robust provided the refugial populations are ancient relative to the age of the species (Figure 3.1).

3.3 Results and Discussion

The search for 'eur* AND phylogeo*' yields 800 papers up to the end of 2006. Including the extra search terms increases this to 820 papers in 164 journals. 439

(53.5%) of these papers are in just ten of the journals and the top ten ranked authors have at least 11 papers each. Of the 820 papers, only 79 meet the criteria of this review. Of these, only 19 identified a putative region of origin for the study taxon, and 60 were unresolved. The unresolved studies included several prominent widespread taxa, including the brown bear *Ursus arctos* (Taberlet & Bouvet 1994; Sommer & Beneke 2005) and dominant forest trees including oaks (Dumolin-Lapegue *et al.* 1997) and common beech *Fagus sylvatica* (Demesure *et al.* 1996; Magri *et al.* 2006). For taxa that are otherwise well-studied, it is perhaps surprising that the geographic origin has remained unresolved. The 19 studies that identified a putative region of origin included 17 whose conclusions supported the 'out of the east' hypothesis, one that supports the 'central European diversification' hypothesis and one that supports the 'out of the west' hypothesis.

3.3.1 Support for the 'out of the east' hypothesis

A total of 17 studies reported an eastern origin (Table 3.1). This conclusion is considered robust according to the criteria defined above in 9 of these studies.

3.3.1.1 Holocene range expansion

Holocene (or estimated time intervals that include the Holocene) range expansion has been reported in eleven species, including fungi (fly agaric *Amanita muscaria* – Geml *et al.* 2006), plants (*Arabis alpina* – Koch *et al.* 2006; perennial rye grass *Lolium perenne* – Balfourier *et al.* 2000), fish (spined loach *Cobitis taenia* – Culling *et al.* 2006; catfish *Siluris glanis* – Triantfyllidis *et al.* 2002), birds (red-crested pochard *Netta ruffina* – Gay *et al.* 2004; willow tit *Parus Montana* – Kvist *et al.* 2001; Pavlova *et al.* 2006) and mammals (moose *Alces alces* – Hundertmark *et al.* 2002; house mouse *Mus musculus* – Boursot *et al.* 1996; Gunduz *et al.* 2005).

Table 3.1. Summary of the 17 studies that support the 'out of the east' hypothesis. Origin refers to the region identified by the authors as the probable geographic origin of the taxon. Dates of most recent (MRCA) and most ancient (MACA) common ancestors of Western Palaearctic populations are given in millions of years before present (mya) and presented as ranges, estimates with confidence intervals, or periods according to the format of the original study. Marker names in bold are non-sequence markers.

Group	Taxon	Common name	Origin	MACA (mya)	MRCA (mya)	Robust†	Markers*	Reference
Fungi	*Amanita muscaria*	Fly Agaric	Beringia	7.5 (3.0-12.0)	Holocene		ITS LSU β-tubulin	Geml *et al.* (2006)
Grasses	*Lolium perenne*	Perennial Rye Grass	Fertile Crescent		0.01		**PCR-RFLP**	Balfourier *et al.* (2000)
Herbs	*Arabis alpina*	Arabis	Caucasus/Middle East	0.5 (0.3-1.4)	0.2 (0.0-1.0)	✓	ITS *trnL-F*	Koch *et al.* (2006)
Insects	*Andricus quercustozae*	Oak Gallwasp	Anatolia		3.2	✓	*cytb* **Allozyme**	Rokas *et al.* (2003)
Fish	*Barbus barbus*	Barbel	Black Sea	2.4-3.5	0.6	✓	*cytb*	Kotlik *et al.* (2004)
	Cobitis taenia	Spined Loach	Black Sea		0.2-0.5		*cytb*	Culling *et al.* (2006)
	Cottus gobio	Bullhead	Black Sea		1.0	✓	*ctr*	Englbrecht *et al.* (2000)
	Leuciscus cephalus	Chub	Tigris-Euphrates basin		2.5-3.0	✓	*cytb*	Durand *et al.* (2000)
	Silurus glanis	Catfish	Ponto-Caspian		Holocene		**μ-sat**	Triantfyllidis *et al.* (2002)
Reptiles	*Emys orbicularis*	Pond Turtle	Eastern	3.2		✓	*cytb*	Lenk *et al.* (1999)
Birds	*Netta ruffina*	Red-crested Pochard	Central Asia		1800s		*ctr1* **μ-sat**	Gay *et al.* (2004)
	Parus montanus	Willow Tit	SE Asia	1.5-2.0	Holocene	✓	*ctr1 ctr2*	Kvist *et al.* (2001)
	Parus spp.	Willow Tit	SE Palaearctic		Holocene	✓	ND2	Pavlova *et al.* (2006)
Mammals	*Alces alces*	Moose	Asia		0.02-0.08		*ctr*	Hundertmark *et al.* (2002)
	Cervus elaphus	Red Deer	Kyrgyzstan-Northern India	7.0		✓	*cytb*	Ludt *et al.* (2004)
	Mus musculus domesticus	House Mouse	Asia (Iraq)		Holocene		**RFLP**	Boursot *et al.* (1996)
	Mus musculus domesticus	House mouse	Asia		Holocene		D-loop	Gunduz *et al.* (2005)

† Studies were considered to provide robust support for the 'out of the east' hypothesis if they met the criteria defined in Section 3.2.2.

* Abbreviations used for sequence markers are as follows: ribosomal internal transcribed spacer (ITS), large subunit (LSU) and D-loop; nuclear β-tubulin; mitochondrial NADH dehydrogenase (ND2), cytochrome *b* (*cytb*) and control region (*ctr*); and chloroplast *trnL-F*.

For two of the eleven taxa described above, corroborating evidence strengthens support for an eastern origin. Willow tits have sibling species in the proposed area of origin (Kvist *et al.* 2001; Pavlova *et al.* 2006), and *Arabis alpina* shows evidence of recent speciation (Koch *et al.* 2006).

3.3.1.2 Pre-Holocene range expansion

All six studies reporting pre-Holocene range expansion provide robust evidence for the inferred geographic origins by our criteria. Pre-Holocene range expansion has been reported in insects (the gallwasp *Andricus quercustozae* – Rokas *et al.* 2003), fish (barbel, *Barbus barbus* – Kotlik *et al.* 2004; bullhead, *Cottus gobio* – Englbrecht *et al.* 2000; chub, *Leuciscus caphalus* – Durand *et al.* 2000), reptiles (European pond turtle, *Emys orbicularis* – Lenk *et al.* 1999) and mammals (red deer, *Cervus elaphus* – Ludt *et al.* 2004).

3.3.2 Support for the 'central European diversification' hypothesis

Only one study in our survey (Ursunbacher *et al.* 2006) provides robust evidence in support of this hypothesis. Mitochondrial cytochrome *b* (*cytb*) and control region (*ctr*) sequence data were used to demonstrate that adders *Vipera berus*, have diverse populations in the Balkan and Italian refugia and a separate diverse refugial clade with a range extending north to Scandinavia and east across the Eastern Palaearctic (Ursunbacher *et al.* 2006). The estimated MRCA date of 1.4 (standard error ± 0.3) mya places this study in the 'pre-Holocene range expansion' category.

3.3.3 Support for the 'out of the west' hypothesis

Only one study in our survey (Jaarola & Searle 2002) provides robust evidence in support of this hypothesis. Sequence data for *cytb* for the field vole *Microtus agrestis* show diversity decreasing from southern (including Spain) through western (including central and northern Europe) to eastern (including northern Europe and Russia) clades (Jaarola & Searle 2002). The estimated MRCA date of 0.5-0.9 (range

of standard errors 0.4-1.1) mya, places this study in the 'pre Holocene range expansion' category.

3.3.4 Studies in which the origin is unresolved

The studies that were placed in this category (see Appendix 1) had aims that did not include longitudinal phylogeography. The discussion below is thus entirely non-judgemental and the data that they present are likely to provide a valuable foundation for future work on taxon origins. The potential for extending previous studies of specific taxa to address longitudinal phylogeography depends largely upon the reasons for the failure of the original study to resolve a geographic origin. These reasons are grouped into three categories, associated with the following issues: (i) rooting; (ii) sampling; and (iii) lack of phylogeographic resolution.

3.3.4.1 Rooting

In 25 studies (Table 3.2), the basal branching order remained unresolved due to the use of an unrootable marker (e.g. common beech *Fagus sylvatica* – Demesure *et al.* 1996) or lack of an appropriate outgroup (e.g. European perch *Perca fluviatilis* – Nesbø *et al.* 1999; marble gallwasp *Andricus kollari* – Stone *et al.* 2001; European hedgehogs *Erinaceous europaeus* and *E. concolor* – Santucci *et al.* 1998).

Table 3.2. Taxonomic distribution of studies in which geographic origin is unresolved. Superscript numbers refer to references listed in Appendix 1.

Group	Studies	Unresolved (%)	Rooting	Sampling	Structure	Total
Fungi	2	*0*				
Ferns	2	*100*			2 [1-2]	**2**
Grasses	2	*50*			1 [3]	**1**
Herbs	2	*50*	1 [4]			**1**
Trees	10	*80*	6 [5-10]	2 [11-12]		**8**
Insects	12	*67*	2 [13-14]	3 [15-17]	3 [18-20]	**8**
Fish	7	*71*	2 [21-22]			**2**
Amphibians	2	*100*	1 [23]	1 [24]		**2**
Reptiles	2	*0*				
Birds	12	*75*		7 [25-31]	2 [32-33]	**9**
Mammals	30	*77*	13 [34-46]	10 [42, 47-55]		**23**
Total	**79**	***71***	**25**	**21**	**8**	**56**

3.3.4.2 Sampling

A further 21 studies (Table 3.2) were unresolved because their sampling design provided data inappropriate for longitudinal phylogeography. Either the species' geographic range extended beyond the sampled area (e.g. for the marble gallwasp *Andricus kollari* – Hayward & Stone 2006), or sampling across the distribution was too sparse in at least one region (e.g. for the bluethroat *Luscinia svecica* – Zink *et al.* 2003). Such studies, nevertheless, represent a potentially valuable resource in longitudinal phylogeography since addition of further samples can balance sample distribution across the range of target taxa (e.g. Stone *et al.* in press).

3.3.4.3 Lack of phylogeographic resolution

In eight studies (Table 3.2), there was a lack of phylogeographic structuring of the chosen marker. In order to detect phylogeographic patterns, there must be an appropriate match between accumulation of divergence in the marker (the mutation rate) and the timescale of the process. In one of these studies, the chosen marker proved too conservative to reveal phylogenetic relationships. Fjelheim *et al.* (2006) sequenced 1940 base pairs of non-coding chloroplast DNA in 56 accessions of meadow fescue, *Festuca pratensis*, and obtained just three haplotypes. Lack of suitable sequence markers is a common problem in phylogeography of plants (Schaal *et al.* 1998). Plant mitochondria do not have a single circle of non-recombining mitochondrial DNA (mtDNA) as is the case for animals. Several different sized circles of mtDNA are present in each plant mitochondrion (Hanson & Folkerts 1992) so paralogues of each marker may be present. The rate of sequence evolution in plant mtDNA is also up to 100-fold slower than animal mtDNA (Palmer & Herbon 1988). Chloroplast DNA (cpDNA) provides an alternative organellar marker in plants. As for mtDNA, inheritance of cpDNA is typically maternal (Hachtel 1980), although inheritance may be paternal (Wagner *et al.* 1987) or biparental (Metzlaff *et al.* 1981) in some cases. There is also evidence for genetic exchange between the chloroplast and nuclear genomes (Baldauf & Palmer 1990), although this is infrequent and does not affect intraspecific studies. The greatest barrier to the use of cpDNA in plant phylogeography is its low mutation rate. Since longitudinal

phylogeography commonly requires less rapidly-evolving markers than latitudinal phylogeography, questions on both timescales should be more commonly addressed as suitable markers are established.

In the small tortoiseshell butterfly, *Aglais urticae*, there was a lack of phylogeographic resolution despite the use of two mitochondrial markers (Vandewoestijne *et al.* 2004). Given that the effective population size of mitochondrial genes is one quarter that for nuclear genes, phylogeographic resolution in such cases is likely to remain difficult to achieve even with additional nuclear sequence data.

3.3.5 Implications for setting conservation priorities

The majority of studies that reported a geographic origin supported the 'out of the east' hypothesis, and this pattern remains under the most conservative criteria for robust results (8 of 10 taxa with robustly supported origins, with 2 Holocene and 6 pre-Holocene westwards range expansions). Eastern populations were most commonly identified as the sources of recent range expansions into Europe (proposed in 11 studies, and a recent eastern origin was robustly supported by our criteria in 2) (Boursot *et al.* 1996; Balfourier *et al.* 2000; Kvist *et al.* 2001; Hundertmark *et al.* 2002; Triantfyllidis *et al.* 2002; Gay *et al.* 2004; Gunduz *et al.* 2005; Culling *et al.* 2006; Geml *et al.* 2006; Koch *et al.* 2006; Pavlova *et al.* 2006). These studies highlight the potential across a range of taxa for rapid colonisation of the entire Western Palaearctic from eastern refugia. In the remaining cases, eastern populations represent genetically divergent source populations for range expansions taking place before or during the Pleistocene (Lenk *et al.* 1999; Durand *et al.* 2000; Englbrecht *et al.* 2000; Rokas *et al.* 2003; Kotlik *et al.* 2004; Ludt *et al.* 2004). The significance of eastern, extra-European, populations over both timescales highlights the importance of incorporating them in the development of conservation strategies.

In the two studies that provided support for the alternative 'central European diversification' (Ursunbacher *et al.* 2006) and 'out of the west' (Jaarola & Searle

2002) hypotheses, genetic diversity was lower for the eastern populations than for the European populations. Eastern diversity was higher for the field vole (Jaarola & Searle 2002) than for the adder (Ursunbacher *et al.* 2006). The field vole was also the only one of the two taxa for which eastern populations formed a monophyletic clade (Jaarola & Searle 2002). These differences in two taxa with similar MRCA dates and non-eastern inferred origins demonstrate the need to consider the full range of each taxon individually to determine the conservation value of eastern populations.

3.4 Conclusions from the literature survey

Of the 79 studies that meet the criteria of this review, most (68) do not provide robust evidence (as defined by the criteria above) for the geographic origin of an extant Western Palaearctic species. Of the 11 studies reporting a robustly supported origin, there is majority support (nine versus two) for the 'out of the east' hypothesis of eastern origin of widespread Western Palaearctic taxa whose ranges extend east of Europe. Longitudinal phylogeography is still in its infancy so it is not yet clear whether the pattern identified in the small number of appropriate studies identified here will be supported by further work in this field. Nevertheless, the presence of a dominant pattern of eastern origin in the literature to date, supported by results from biogeographic studies that show an 'out of Asia' pattern for European taxa (Sanmartin *et al.* 2001; Donoghue & Smith 2004), suggests that there is value in routinely incorporating eastern populations into phylogeographic studies. This review has highlighted the potential for using existing markers and methods to address phylogeographic questions from a longitudinal perspective. The list of unresolved studies presented in Table 3.2 and Appendix 1 provides a summary of potential target taxa from the latitudinal phylogeography literature. Widespread adoption of a longitudinal perspective (using the protocol described above) will provide valuable information on the processes underlying current refugial distributions, which have historically been viewed as the starting point for phylogeographic studies. Informed decisions on conservation priorities require objective delimitation of taxa, which can often only be achieved by thorough

longitudinal sampling that includes areas that have historically been less well studied. Understanding of relationships between refugial populations and geographic origins through longitudinal phylogeography across a greater range of taxa will provide information on the role of areas to the east of Europe as both cradles and museums of Western Palaearctic diversity. By assessing concordance in patterns across species, longitudinal phylogeography may be used to distinguish between alternative models of community development. Communities may be found to have shared longitudinal histories, rather than developing by local recruitment from existing pools of taxa following the dispersal of a key taxon. Under this scenario, the importance of conservation priorities would need to be set at the ecosystem rather than single taxon level.

3.5 A suggested protocol for longitudinal phylogeography

The literature survey has demonstrated the potential for shared patterns in longitudinal phylogeography, but also highlighted the paucity of published studies to date that address the specific aims of longitudinal phylogeography to reach robustly supported conclusions. The aim of this section is to provide a review of suitable methodology for longitudinal phylogeography to act as a basis for the design of future studies.

3.5.1 Suitable taxa

The requirement for relatively ancient MACAs for each population may be relaxed if there is corroborating evidence to suggest that the current distribution of diversity reflects the phylogeographic history of the taxon. Previous phylogeographic studies (as identified in Table 3.2 and Appendix 1) may be used to guide the choice of appropriate taxa and, since the same markers may be used in longitudinal phylogeography, existing datasets can be built upon.

3.5.2 Molecular markers

Patterns accompanying range expansion depend on timescale and the evolutionary rate, the information content and the effective population size of the marker (Roderick 1996; Section 2.1). Glacial refuges that have been separated without migrants for 10,000 years may be in the process of sorting ancestral polymorphism for allozyme loci (high population size, low mutation rate), but have mutually discrete clades of mitochondrial haplotypes (one quarter of the population size for nuclear loci, higher mutation rate) (e.g. Stone *et al.* 2001, Stone *et al.* 2007). Populations that show little or no divergence in either of these markers may still be differentiated using sequence data for long, highly variable intron sequences (e.g. Davies *et al.* 1999). Selection of a marker showing adequate variation among possible sources allows the origin of invading individuals to be identified using source-specific polymorphism (e.g. Davies *et al.* 1999, Demesure *et al.* 1996, Dumolin-Lapègue *et al.* 1997, Ferris *et al.* 1995, Geller *et al.*, 1997, Konnert & Bergmann 1995, Matyas & Sperisen 2001, Stone *et al.* 2001, Stone *et al.* 2007). In Europe, a sufficient number of species' colonisation routes have been identified in this way to allow major latitudinal invasion routes to be identified (Hewitt 1999, Taberlet *et al.* 1998). An alternative for species showing low genetic diversity is to study genetic patterns in closely associated organisms, such as commensals or parasites: for example, human range expansion has been reconstructed using genetic diversity in disease pathogens in the neotropics (Fisher *et al.* 2001) and in commensal rats in Polynesia (Matisoo-Smith *et al.* 1998).

The ideal data for longitudinal phylogeography are sequences for markers that are evolving at an appropriate rate from several individuals from each population. Sequence data allow detailed inference of the relationships between populations as well as reconstruction of derived and ancestral states and, with the testable assumption of a molecular clock, dates of divergence. Although sequence data should usually be preferred to other types of evidence for longitudinal phylogeography, it is important to consider potential problems with inferences based on sequence data. Individual markers may not be representative of the species'

history, especially for animal mitochondrial genes whose phylogeographic signatures can contain artefacts associated with introgression and/or selective sweeps (associated with *Wolbachia* and other microbial symbionts; Hurst & Jiggins 2005). For a gene to be representative of a species history it must be evolving neutrally, and selection is likely to generate gene trees whose structures do not represent species phylogeographic history. As with all levels of phylogenetics, datasets for multiple markers (including both nuclear and organellar DNA) are thus preferable. The assumption of selective neutrality should be tested for each marker using, for example, the McDonald-Kreitman Test (McDonald & Krietman 1991) before proceeding with phylogeographic reconstruction. Despite reduced costs associated with large-scale sequencing projects, there are still relatively few candidate markers for longitudinal phylogeography, particularly for plants. Commonly used markers that have been applied on the timescale of longitudinal phylogeography (see Table 3.1) include mitochondrial genes in animals (*ctr*; *cytb*; and NADH dehydrogenase, ND2), ribosomal (internal transcribed spacer, ITS; and D-loop) and other nuclear sequences in animals and plants (β-tubulin) and some chloroplast markers in plants (e.g. *trnL-F*).

3.5.3 Sampling regime

It is essential that the entire range of the target taxon should be sampled as evenly as possible. Increased sampling effort in specific areas may bias phylogeographic reconstruction, particularly if reconstruction is based on the diversity of unrooted markers. It is important to ensure that the phylogeographic analysis is of a monophyletic lineage for the longitudinal approach to be valid. Since this latter point is usually unknown prior to extensive longitudinal sampling (see Challis *et al.* 2007 for a cautionary example; Chapter 4), all potential sibling taxa should be included in the study design.

3.5.4 Choosing the analytical approach for sequence data

Reconstruction of phylogeographic relationships is, arguably, the most important and most controversial aspect of phylogeography. Two classes of methods are used (Section 2.7): (i) phylogenetic trees, which assume a single evolutionary history for an entire sequence alignment; and (ii) phylogenetic networks, which are reticulated networks designed to accommodate events such as recombination or horizontal gene transfer. The principal arguments in favour of trees are the ability to use realistically complex evolutionary models (Section 2.3) and the ability to exploit powerful techniques for tree analysis, such as trait mapping (Section 2.9) and node dating (Section 2.8). Phylogenetic tree reconstruction using Bayesian Inference (BI; Section 2.2.4) allows simultaneous estimation of tree topology and node support while integrating over uncertainty in the data. A powerful extension of this has been the application of Bayesian model averaging (BMA; Section 2.3.4) to integrate over uncertainty in the choice of evolutionary model (Nylander 2004).

The principal arguments in favour of networks are the ability to account for recombination of markers and genetic exchange between populations, which may be frequent in the absence of reproductive isolation (Section 2.7.1). The argument in favour of using networks to account for exchange of genetic material when reconstructing intraspecific relationships is particularly strong for nuclear markers (which are likely to be affected by recombination), but much weaker for mitochondrial data (due to maternal inheritance and lack of recombination). However, when samples are collected on the basis of phenotype (which is essentially nuclearly determined), introgression will also generate reticulation in mitochondrial sequence data. Both approaches allow haplotypes to be treated as internal nodes, whether through the explicit assumption that intermediate haplotypes are still present in a population for recent events (statistical parsimony) or implicitly, through allowing tips to have zero branch length.

Most phylogeographic studies that use networks employ either median-joining or parsimony networks. However, the complexity of evolutionary models

supported in other types of phylogenetic networks (Huber *et al.* 2002; Bryant & Moulton 2004) is beginning to rival those available for phylogenetic tree reconstruction (Section 2.7.2) so the debate over trees versus networks for phylogeography must be about the true nature of intraspecific data and not the limitations of some of the widely used network methods.

Although the choice between tree and network methods is often a matter of personal preference, a formal test for the tree-likeness of molecular data has been proposed by Huson & Bryant (2006). This test assesses the extent to which the data are compatible with a bifurcating tree in a two step process: (i) a 95% confidence phylogenetic network is generated in SplitsTree (Huson & Bryant 2006); then (ii) the hypothesis that the data originated on a tree is accepted if the 95% confidence network contains a tree and rejected if it does not (Section 2.7.3). Full consideration of the relative merits of the two approaches is beyond the scope of this review, so protocols for analysis of both trees and networks are presented below.

3.5.5 Node dating

In order to calculate absolute dates for nodes that represent MRCAs and MACAs (Section 2.8.3) of a taxon or of geographic clades, the hypothesis of molecular clock-like evolution must first be tested, and if clock-like evolution is supported, its evolutionary rate must then be calibrated (Pulquério & Nichols 2007). An alternative, but much less desirable, approach is the use of generic calibrations, such as the 2% pairwise divergence per million years for invertebrate mitochondrial DNA (Brower 1994). Any estimate of absolute node ages can only be as accurate as the calibration rate used (Pulquério & Nichols 2007) so estimates based on generic calibration include an extra degree of uncertainty and must be treated conservatively (Section 2.8.4). In the absence of acceptable data for calibration, uncalibrated, relative node ages in mutational units can still be compared and may be informative (e.g. Hayward & Stone 2006).

3.5.5.1 Trees

The hypothesis of a strict molecular clock can be tested on a model-based phylogenetic tree by performing the phylogenetic reconstruction both with and without a constant evolutionary rate throughout the tree (molecular clock; Section 2.8.1). Confidence intervals around molecular clock estimates follow a Poisson distribution and can become very large when reconstructing deep nodes. Date estimates based on a molecular clock should ideally be checked against the fossil record to reduce this uncertainty and confirm that the results are credible (Pulquério & Nichols 2007). For many taxa of interest for longitudinal phylogeography, such confirmation will not be possible.

3.5.5.2 Networks

There are currently no network-specific molecular clock methods available. Studies using networks typically calculate MRCA and MACA dates from sequence divergence estimates for geographic clades (Hoelzer *et al.* 1998), effectively assuming a strict molecular clock (e.g. Rokas *et al.* 2003; Kotlik *et al.* 2004).

3.5.6 Geographic reconstruction

3.5.6.1 Trees

Reconstruction of ancestral distributions can be achieved by mapping location onto a phylogeny to infer the distributions of MRCAs of taxa and clades. Parsimony (Section 2.2.1), maximum likelihood (ML; Section 2.2.3) and BI (Section 2.2.4) methods are available, the latter two can be applied to the full set of trees sampled during BI phylogeny reconstruction using the program BayesMultiState (Section 2.12.2). The advantage to model-based trait mapping is that historic dispersal routes can be inferred using model simplification to determine a minimum number of supported geographic transitions. The protocol is described below for ML (using likelihood-ratio tests, LRTs; Section 2.3.3.3), but this could be substituted for BI (with Bayes Factors, BF; Section 2.3.3.4) if required. Using BF, the distinction

between nested and non-nested comparisons is unnecessary. An illustrated example is presented in Figure 3.2.

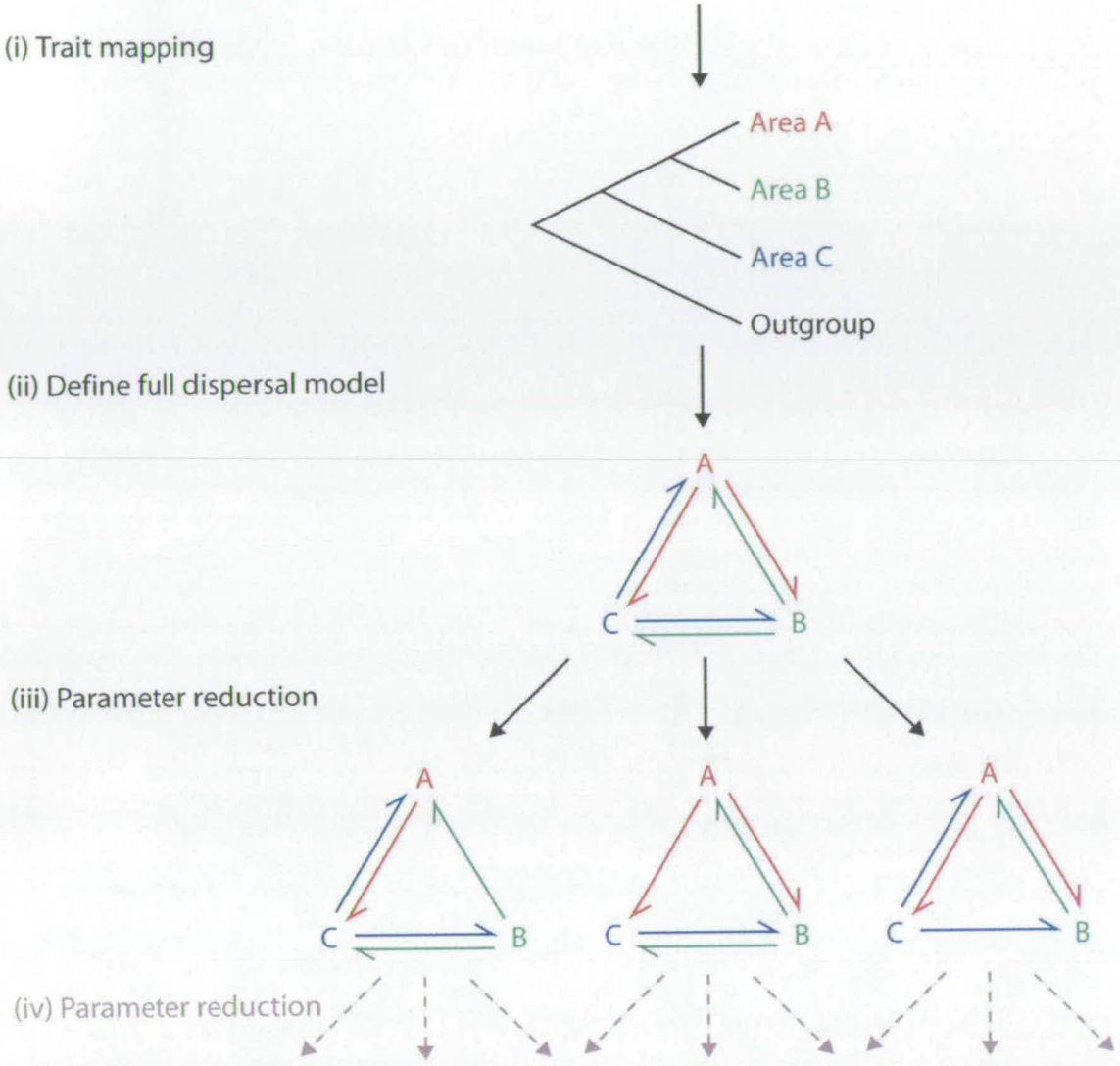


Figure 3.2 Determining geographic origin and dispersal histories by model-based trait mapping with parameter reduction. In this hypothetical example, one individual of a taxon was collected from each of three geographic areas. The area of each sample was mapped onto a phylogeny (i) and character state changes (dispersal events) and ancestral state probabilities were estimated under the full model illustrated above (ii) using BayesMultiState (Pagel *et al.* 2004). By comparing the likelihood obtained under each of the reduced models (iii; only a subset of possible models are illustrated) with the likelihood obtained under the full model using likelihood ratio tests, the importance of the parameter in describing the dispersal history of the taxon can be assessed. Parameters that can be removed without significantly affecting likelihood reflect dispersal routes that are unlikely to have been used by sufficient numbers of individuals to have left a phylogeographic signature. Models can be further reduced (iv) until the most likely model (or set of models) has been identified. If all dispersal parameters from one of the geographic areas can be removed without making the model likelihood significantly worse, the area concerned is not likely to represent the geographic origin.

In principle, all transitions between alternative geographic regions are possible during the phylogeographic history of a taxon (the full, unordered model). However, it is unlikely that all possible transitions actually occurred, and estimation of redundant parameters creates an unnecessarily complex model. Once an initial likelihood is obtained under the full, unordered model (bidirectional changes possible between all character states), systematic removal of unidirectional exchange parameters allows alternative (reduced) models to be compared. Each reduced model is nested within the full, unordered model so reduced models can be compared with the full unordered model using hierarchical LRTs (hLRTs; Frati *et al.* 1997; Huelsebeck & Crandall 1997) by comparing the ratio of likelihoods (difference between log-likelihoods) against the χ^2 distribution (Goldman 1993). It is likely that a number of reduced models will be significantly better than the full, unordered model. Since comparisons between unordered models are not necessarily nested, it may be necessary to use the rule of thumb that two log-likelihoods constitutes a significant difference between competing hypotheses (Edwards 1972; Pagel 1999). Once all comparisons have been made, the best supported model (or set of models) will indicate the probable directions and frequencies of phylogeographic dispersal events.

3.5.6.2 Networks

Trait mapping across networks is less straightforward as networks must first be rooted to form a directed acyclic graph (DAG). Likelihood procedures can be applied to root networks, attaching the root to produce the most likely DAG (Strimmer & Moulton 2000), which may provide a means of inferring geographic origin. An alternative approach, Nested Clade Analysis (NCPA; Templeton *et al.* 1995; Section 2.7.2), has been widely adopted in latitudinal phylogeographic studies (Templeton 2004). NCPA provides a statistical framework for assigning historic processes to describe the geographic distribution of haplotypes. This could potentially be extended to include longitudinal phylogeography questions but not inference of geographic origin.

3.5.7 Troubleshooting

3.5.7.1 Improving basal resolution

The mitochondrial gene sequences that are widely-used in phylogeography, including *coxI* (cytochrome oxidase I), *cytb* and *ctr* (Hewitt 2004), provide good resolution of relatively recent processes (i.e. mid-Pleistocene – Holocene). Over longer timescales, however, they are vulnerable to substitution saturation (Lopez *et al.* 1999; Philippe & Forterre 1999, Rokas *et al.* 2002), which may reduce the phylogenetic resolution at the basal nodes that are of interest in longitudinal phylogeography. In studies using sequence data with an outgroup, the use (if possible) of more closely related outgroups may improve basal resolution. If this approach is unsuccessful, or there is no appropriate extant outgroup, it may be necessary to introduce an additional marker whose mutation rate is appropriate to the timescale of the longitudinal range expansion process. In the case of rapid dispersal between refugia during the early- or pre-Pleistocene, no sequence marker is likely to have acquired sufficient mutations during dispersal to provide resolution. In such situations, patterns in the diversity of non-sequence markers, while sometimes unreliable for the reasons stated above, may provide resolution.

3.5.7.2 Detecting localised extinctions

Where it is otherwise unclear whether a robust interpretation of the geographic origin of a taxon can be made due to the possibility of extinction of basal lineages, lineages-through-time plots may be useful (Nee 2001). Such plots require at least an uncalibrated molecular clock to produce a plot of the number of lineages against the relative time depth of nodes. Under a constant birth rate model, with no lineage extinction, a straight line is expected with gradient equal to the rate of lineage diversification (Nee 2001). Departures from this expectation can result from actual processes or may be artefactual (Barraclough & Nee 2001). An ancient upturn in gradient would suggest the possibility of basal lineage extinction and the data are likely to be unsuitable for longitudinal phylogeography.

Chapter 4

Longitudinal range expansion and cryptic eastern species in the Western Palaearctic oak gallwasp *Andricus coriarius*†

Chapter 3 showed that in the limited number of relevant studies published to date, there are dominant patterns of eastern origin for Western Palaearctic taxa but highlighted the need for further research in this area. This chapter presents nuclear and mitochondrial sequence evidence from the entire geographic range of Andricus coriarius *to investigate the genetic legacy of longitudinal range expansion.*

4.1 Introduction

Andricus coriarius belongs to the host-alternating *Andricus* clade (Section 1.3.5) and is one of at least ten host-alternators with a longitudinal distribution extending from Iberia and Morocco to Iran and the Caucasus (Atkinson 2000; Melika *et al.* 2000; Nieves-Aldrey 2001; Cook *et al.* 2002; Rokas *et al.* 2003a,b). Unlike other host-alternators, the distribution of *A. coriarius* has remained largely unaffected by the human-mediated dispersal of section *Cerris* oaks into northern Europe (Section 1.2.5) and has remained restricted to the latitudinal band in which the native ranges of section *Cerris* and section *Quercus s.s.* oaks overlap (Section 1.2.3). Phylogeographic patterns in *A. coriarius* should therefore reflect natural responses to the longitudinal changes in section *Cerris* species distributions (Section 1.3.7).

A longitudinal approach to oak gallwasp phylogeography raises three general questions: (a) which region within these wide distributions was the 'cradle' for these species, (b) what is the timescale and mode (continuous or iterated) of longitudinal range expansion, and (c) what is the frequency and direction of shifts between alternative oak hosts? Phylogeographic analysis of another widespread host-alternating *Andricus* species, *A. quercustozae*, supported an eastern origin (Rokas *et al.* 2003a), leading to the proposal of an 'Out of Anatolia' hypothesis for host-

† A modified version of this chapter has been published as:
Challis RJ, Stone GN, Mutun S, Nieves-Aldrey J-L, Preuss S, Rokas A, Aebi A, Sadeghi E & Tavakoli M (2007) Longitudinal range expansion and cryptic eastern species in the Western Palaearctic oak gallwasp *Andricus coriarius*. *Molecular Ecology*, **16**, 2103-2114.

alternating *Andricus* (Section 1.3.6). This pattern of eastern origin was reflected in the analysis of a diverse set of Western Palaearctic taxa (Chapter 3).

The aim of this chapter is to extend the longitudinal range sampled in previous studies of oak gallwasps to Lebanon and Iran, and so include the full known range of host-alternating *Andricus* in the Western Palaearctic. Analysis of sequence data from 80 individuals of *A. coriarius* for a 433 base-pair (bp) fragment of the mitochondrial cytochrome *b* gene, with additional sequence data for the D2 region of the nuclear 28S region, is used to detect cryptic lineages and assess their phylogenetic status. This chapter also assesses the demographic processes associated with longitudinal range expansion, to investigate whether natural longitudinal range expansion has left a signature of rapid population growth, as suggested in other species (e.g. Culling *et al.* 2006, Michaux *et al.* 2003).

4.2 Materials and Methods

4.2.1 Sample collection

Parthenogenetic generation galls of *A. coriarius* were collected from locations across the species' range from Spain to Iran (Appendix 2; Figure 4.1). All galls were morphologically similar: galls were globular, with a surface coating of stout pointed spines, contained many larval chambers, and developed on lateral and terminal buds of young shoots. Among recognised Western Palaearctic species, this combination of traits is only present in *A. coriarius*. As far as possible, sample sizes were balanced across distribution regions, with additional sampling effort in Iran after preliminary investigation revealed high haplotype diversity. The only sexual generation oak host available to the sampled Iranian populations was *Q. libani*, while both *Q. libani* and *Q. cerris* were available to Lebanese populations. All populations further west have sexual generations on *Q. cerris*, with the exception of Iberian populations, for which *Q. suber* is the only available sexual generation host. The parthenogenetic females that emerge from a single gall are commonly the offspring

of a single sexual female (Atkinson *et al.* 2002). Only a single individual was thus sequenced from each gall.

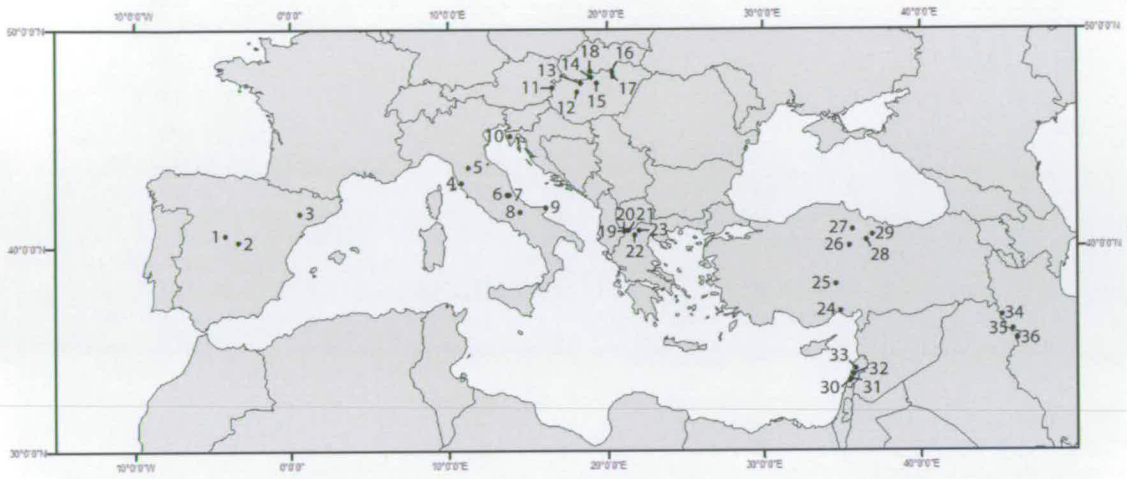


Figure 4.1 Map showing the collection locations. Locations are numbered as in Appendix 2.

4.2.2 DNA extraction and sequencing

DNA was extracted using the DNeasy Tissue kit (Section 2.11.2). 433 bp fragments of the mitochondrial cytochrome *b* (*cytb*) gene were amplified (Section 2.11.3.1) and sequenced (Section 2.11.4) for 77 individuals. A subset of individuals was also sequenced for the D2 region of the 28S ribosomal array (28SD2; Section 2.11.3.3). All sequences were deposited in Genbank (accession numbers EF029996-EF030036, *cytb*, and EF030037-EF030039, 28SD2). Three previously published *A. coriarius cytb* sequences were also included: Gödöllö, Hungary (AJ228458); Edessa, Greece (AF539556); and El Escorial, Spain (AF539557) (Rokas et al 2003b).

4.2.3 Testing the validity of a phylogenetic framework

The test for tree-likeness (Huson & Bryant, 2006; Section 2.7.3) was applied in SplitsTree 4.4 (Section 2.12.13) to test for tree-like evolution at both the inter- and intra-specific levels using the two-gene sequence alignment The hypothesis of tree-like evolution was accepted for the interspecific relationships, however, intraspecific relationships between *A. coriarius* haplotypes were not compatible with any tree.

Phylogenetic methods have therefore only been used at the interspecific level and alternative network-based methods are adopted for intraspecific analyses.

4.2.4 Phylogenetic analysis

Bayesian phylogenetic reconstruction was performed using MrBayes 3.1 (Section 2.12.10). The two-gene dataset was partitioned by gene and the *cytb* partition was further divided into the three codon positions. Instead of *a priori* model selection, a parameter-rich GTR+I+Γ model was applied to each partition, exploiting the efficiency of the MC^3 algorithm within MrBayes to optimise parameters. Two independent runs of four Markov chains over 2 million generations were performed for each analysis with the temperature parameter set at 0.15. Convergence was assessed through examination of plots of chain parameters and comparison of the output from the two independent runs (Section 2.5). Trees were sampled every 1000 generations after a burn in period of 200,000 generations. To assess the monophyly of *A. coriarius*, additional *cytb* and 28S D2 sequences were included in the phylogenetic analyses. All *cytb* outgroup sequences are previously published: *A. caputmedusae*, AF539553 and *A. conificus*, AF539555 (Rokas et al. 2003b); *A. conglomeratus*, AJ228468, *A. curvator*, AJ228453, *A. seckendorffi*, AJ228449, *A. solitarius*, AJ228475 and *Cynips quercus*, AJ228478 (Stone and Cook 1998); and *A. kollari*, AF242739 (Stone *et al.* 2001). Two previously published 28SD2 sequences were included: *A. curvator*, AF395155 and *A. kollari*, AF395156 (Rokas *et al.* 2002). 28SD2 sequences for *A. caputmedusae*, EF030040, *A. conglomeratus*, EF030041, *A. conificus*, EF030042, *A. seckondorfii*, EF030043, *A. solitarius*, EF030044, and *C. quercus*, EF030045, were provided by Antonio Hernandez-Lopez (unpublished data).

4.2.5 Estimating the time depth of nodes in the haplotype tree

The validity of the assumption of a molecular clock for the *cytb* data was tested using Bayes Factors (Section 2.8.1). Marginal likelihood was approximated by calculating the harmonic mean, *H*, log-likelihood, *lnL* scores obtained from MC^3 runs under a non-clock and three strict clock models (uniform, birth-death and coalescence),

implemented in MrBayes 3.1 (Section 2.12.10). The differences in *H lnL* for each model were interpreted using the values presented by Kass & Raftery (1995). Dates were estimated for both most recent common ancestors (MRCAs) and most ancient common ancestors (MACAs; Section 2.8.3) at all nodes in the Bayesian phylogeny using BEAST (Section 2.12.3). Rates were kept constant across all branches as this was supported by the test for clock-like evolution. The monophyly of MRCAs was not constrained as all reconstructed nodes had 100% posterior probability. The MCMC chain was run for 10 million generations under the GTR model with gamma-distributed rate heterogeneity and sampled every 1,000 generations after a burn-in of 1 million generations. All parameters and likelihood values were assessed using Tracer 1.3 (Rambaut & Drummond 2004) to estimate 95% confidence intervals and ensure a sufficient effective sample size for each parameter.

4.2.6 Nested clade phylogeographic analysis

A nested clade phylogeographic analysis (NCPA; 2.7.2) was performed to reveal the ecological and evolutionary factors likely to have lead to the present phylogeographic distribution (Templeton 1998). A haplotype network was used to define a hierarchical set of nesting clades according to the nesting rules of Templeton *et al.* (1987) and Templeton & Sing (1993). The haplotype network was generated using statistical parsimony in TCS 1.2.1 (Section 2.12.15) with the connection limit set at 95%. Loops were excluded according to the coalescent criteria of Crandall & Templeton (1993) with the geographic criteria of Pfenninger & Posada (2002) applied where this was ambiguous. Clade distances (D_c) and nested clade distances (D_n), calculated from the frequency and locations of haplotypes, were tested against the null hypothesis of random haplotype distribution using GEODIS 2.4 (Section 2.12.7). Where the null hypothesis was rejected, the cause was inferred using the latest version of the key of Templeton *et al.* (1995; available at http://inbio.byu.edu/Faculty/kac/crandall_lab/geodis.htm).

4.2.7 Inference of population demographic history

Pairwise mismatch plots (Section 2.12.1.2 and 2.12.5.3 of substitutional differences between pairs of sequences were calculated and compared against Poisson models under assumptions of constant population size (Slatkin & Hudson 1991; Rogers & Harpending 1992) and population growth-decline (Rogers & Harpending 1992) using DnaSP 4.10 (2.12.5). To meet the assumptions of this test, it was applied only to a subset of haplotypes inferred by the NCPA to represent a single contiguous, expanding population. The raggedness statistic, *r* (Harpending 1994) and Fu's *FS* (Fu 1997; Section 2.12.5.1) were calculated to assess the smoothness of the pairwise mismatch plots.

4.3 Results

4.3.1 Cryptic polyphyly in *A. coriarius*

Single gene phylogenies for *cytb* and 28SD2 each supported the division of *A. coriarius* into three clades, containing haplotypes from: (i) across the sampled range, (ii) Iran only, and (iii) Lebanon only. Alone, the *cytb* data resolved relationships within clades but rooting of each clade within the wider phylogeny of *Andricus* was unclear. The 28SD2 clade provided resolution at the intraspecific level but there was no sequence divergence within the three clades. The following analyses thus use the partitioned two-gene dataset, as described in the Methods.

Bayesian phylogeny reconstruction of the combined cytb/28SD2 dataset showed that rather than representing a single monophyletic lineage, the sampled A. coriarius haplotypes fall into three distinct lineages (Figure 4.2). One lineage (the Lebanese Clade, Figure 4.2) contains only Lebanese haplotypes (for four of the five Lebanese specimens, representing three of the four Lebanese sites). The second (the Iranian Clade, Figure 4.2) contains only Iranian haplotypes (for 15 of the 29 Iranian specimens, representing two of the three sites). The third (the Main Clade, Figure 4.2) contains all remaining haplotypes, including haplotypes from one Lebanese

individual, all 14 individuals from the Iranian site of Baneh, and all European sequences. No single site contained haplotypes in more than one of the three clades.

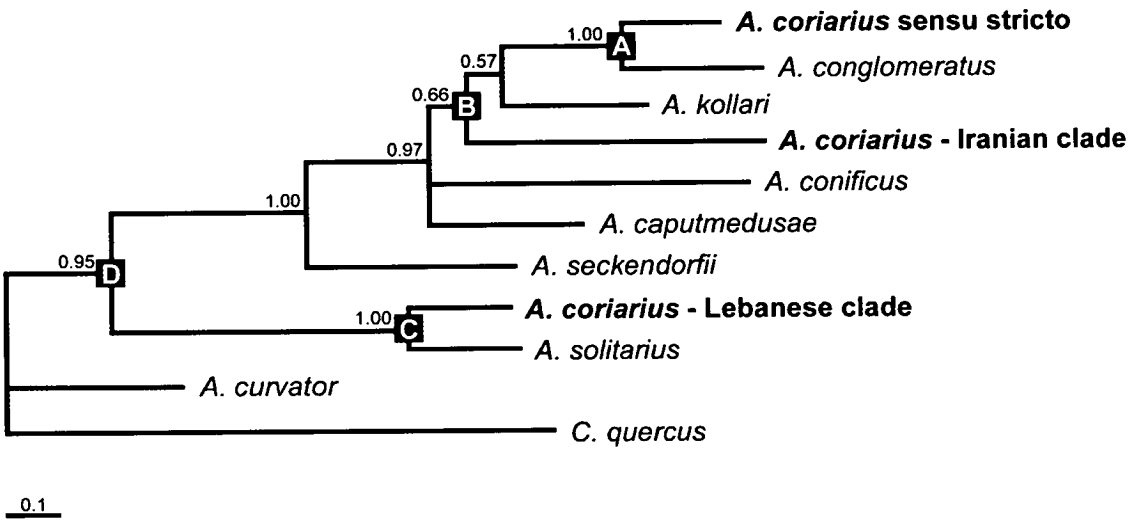


Figure 4.2 Simplified two-gene Bayesian phylogeny showing the locations of the three clades of *A. coriarius* within *Andricus*. Data were partitioned by gene (28SD2 and cytochrome *b*) with the cytochrome *b* gene further partitioned by codon position. A GTR+I+Γ model of evolution was applied to each partition. Two independent MC^3 runs of 2×10^6 generations were sampled every 1000 generations following a burn in period of 2×10^5 generations. Posterior probability support values are shown for each node. Labels A – D indicate nodes for which dates have been estimated (Table 4.1).

Intraspecific relationships within the Main Clade are shown in a statistical parsimony reconstruction of the haplotype network (Figure 4.3). The divergent Lebanese and Iranian haplotype groups could not be connected to the rest of the network (Figure 4.3) at either the 95% or the 90% confidence levels. The Main Clade forms a monophyletic group with the other representatives of the *A. kollari* Clade. The Iranian Clade lies between this clade and representatives of the *A. hartigi* and A. quercuscalicis Clades, although the posterior probability for placing the Iranian Clade outside of the *A. kollari* Clade is very low. The Lebanese Clade is most closely related to *A. solitarius*, a representative of the non-host-alternating *A. fecundator* Clade. *Andricus coriarius* is thus either a paraphyletic or polyphyletic taxon.

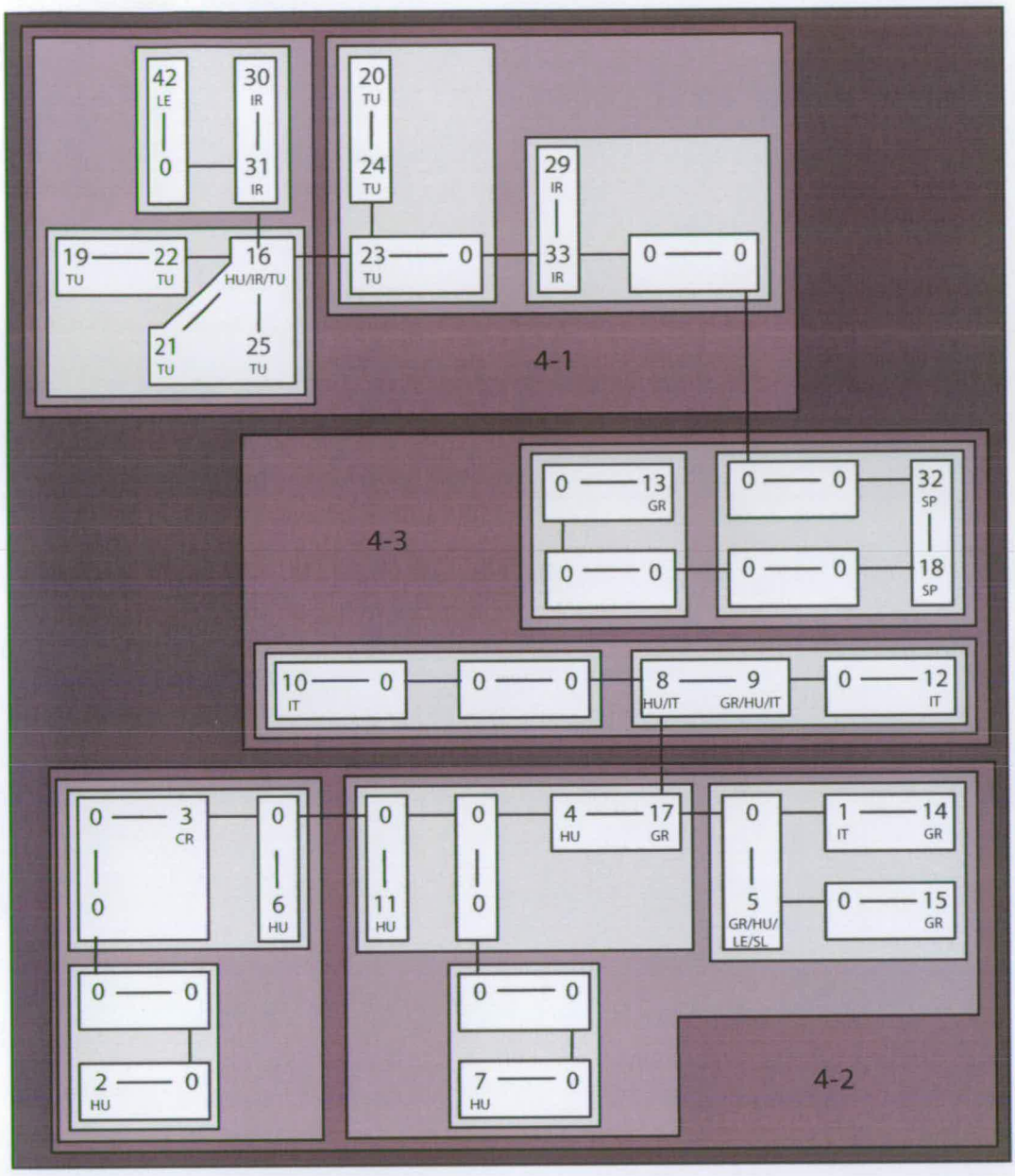


Figure 4.3 Haplotype network for the *A. coriarius* sensu stricto clade estimated using statistical parsimony (95% connection level). Haplotype numbers are as in Appendix 2, and locations are indicated by letters in parentheses (CR, Croatia; GR, Greece; HU, Hungary; IR, Iran; IT, Italy; SL, Slovakia; SP, Spain; TU, Turkey). 0 indicates a missing haplotype.

4.3.2 Dating lineage divergence and longitudinal range expansion in *A. coriarius*

Bayes factor (BF) comparison with the non-clock model (*H lnL* = 1316.19) for the cytb dataset provides strong support for the use of a coalescent clock (*H lnL* = 1310.00; BF = –12.38) but does not support either the uniform or the birth-death

models (BF: 23.92 and 77.60, respectively). Calculation of MRCA/MACA dates in BEAST (Section 2.12.3) was therefore performed under a coalescent model of population growth. Estimated ages of MRCAs and MACAs for key nodes are shown in Table 4.1.

Table 4.1 Ages of MRCAs and MACAs calculated using BEAST (Section 2.12.3). The MRCA for a given clade incorporates sequences only for that clade, while the age of the MACA is calculated by including the nearest sister taxon. Values in parentheses are 95% confidence intervals assuming a clock calibration of 2.3% pairwise divergence per million years.

		Estimated date (mya)	
Node	**Ancestor of**	**MRCA**	**MACA**
A	Main *A. coriarius* Clade	1.6 (1.0 – 2.2)	3.5 (2.5 – 4.8)
B	Iranian Clade	0.82 (0.25 – 1.5)	4.5 (3.1 – 6.1)
C	Lebanese Clade	0.30 (0.037 – 0.66)	3.3 (2.3 – 4.5)
D	All *A. coriarius* Clades	8.9 (5.9 – 12.4)	

The Main *A. coriarius* lineage is inferred to have diverged from the rest of the kollari clade between the mid-Pliocene (MACA 3.5 mya) and the Pleistocene (MRCA 1.6 mya). The Iranian and Lebanese Clades are inferred to have diverged from the Main *A. coriarius* lineage long before the Pleistocene around 4.5 mya (Pliocene) and 8.9 mya (late Miocene), respectively. Of the outgroups in this study, the Lebanese clade clusters closest to *Andricus solitarius*.

Although the divergence between lineages is inferred to be relatively ancient, sequence divergence and nucleotide diversity within clades (Main Clade, 3.6%, π = 0.01554; Iranian Clade, 2.1%, π = 0.00529; and Lebanese Clade, 1.0%, π = 0.00644) are much lower than sequence divergence and nucleotide diversity between clades (Main vs. Iranian Clade, 8.3%, π = 0.03427; Main vs. Lebanese Clade, 11.3%, π = 0.02884; and Iranian vs. Lebanese Clade, 11.5%, π = 0.04742). This contrast within all three species is compatible with population bottlenecks since species divergence.

4.3.3 Nested clade and pairwise mismatch analyses of the Main *A. coriarius* clade

Due to the high sequence divergence between the three lineages of *A. coriarius*, and the associated difficulty of linking these in a single network, NCPA was attempted only for the Main *A. coriarius* clade, using the *cytb* sequence data. The nesting structure for NCPA is shown in Figure 4.3. The highest root probability (11.1%) was assigned to haplotype 16, which was sampled from Hungary, Iran and Turkey. Considering all samples from each country, the greatest mean root probability (5.3%) was shared by Iran and Turkey, suggesting an eastern origin for this clade. All other countries had mean root probabilities below 3.9%. The NCPA revealed statistically significant geographical associations at the fourth nesting level and at the total cladogram level. Clade 4-3 shows contiguous range expansion while at the total cladogram level there is support for historic gradual range expansion with subsequent fragmentation or long distance dispersal. Support for an eastern root suggests that range expansion was predominantly westwards. While it is not possible to distinguish between the two eastern oak species *Q. cerris* and *Q. libani* as ancestral sexual generation hosts, the association with the Iberian host *Q. suber* is probably derived.

A pairwise mismatch plot (Figure 4.4) of the Main *A. coriarius* clade has a single, smooth peak ($r = 0.0093$, $P < 0.05$), which indicates that westwards range expansion was associated with population growth by a single demographic entity. Further test statistics (Fu's $FS = -24.01$, $P < 0.05$) also indicate the departure from neutrality expected in an expanding population. The same distribution is supported in mismatch plots of the fourth nesting level subclades (not shown). Smoothly unimodal mismatch distributions are incompatible with significant population fragmentation, so it is likely that long distance dispersal accounts for the substructure within the Main *A. coriarius* clade.

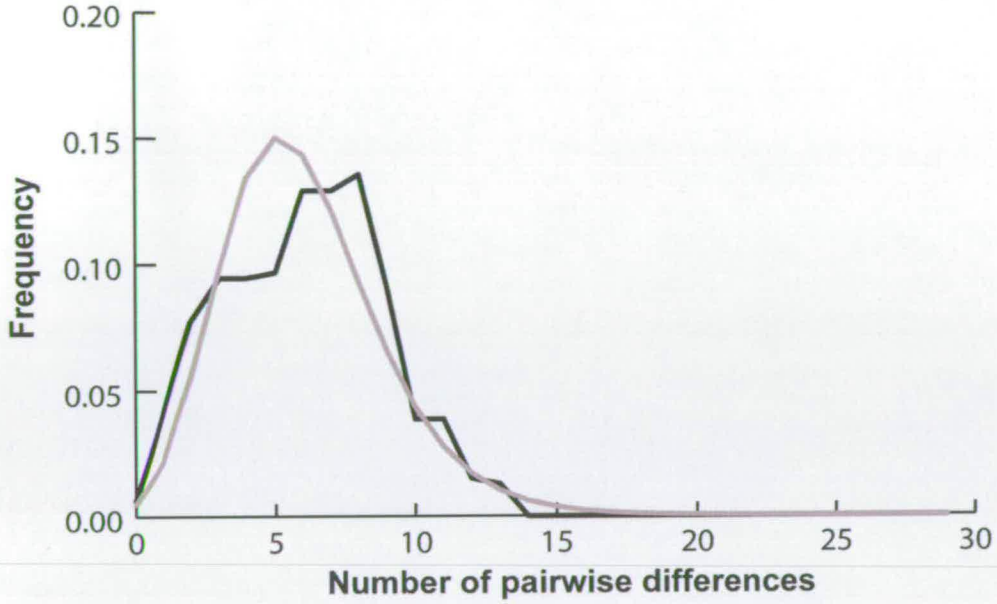


Figure 4.4 Pairwise mismatch plot of the main *A. coriarius* clade. Filled circles indicate observed data. Open triangles indicate the distribution expected under a population growth-decline model (θ initial = 1.414, θ final = 1000, τ = 4.664).

4.4 Discussion

4.4.1 Cryptic lineages in *A. coriarius*

The data show that insects reared from galls currently recognised as representing a single species in fact represent three deeply divergent lineages. The main *A. coriarius* clade contains haplotypes from all sampled countries, and is either locally or regionally sympatric with divergent monophyletic lineages from each of Iran and Lebanon. The Lebanese, and possibly Iranian, Clades lie outside the monophyletic group formed by the main *A. coriarius* clade and its sister taxa in the *A. kollari* clade, and *A. coriarius* (as represented by the sampled individuals) is thus paraphyletic.

Andricus coriarius was described from central Europe by Hartig in 1843 (Hartig 1843). All previous studies of gallwasps from Turkey (Dalla Torre & Kieffer 1910) eastwards through to Iran (Chodjai 1980) and the Caucasus (Maisuradze 1962) have identified insects emerging from the characteristic spiny and multichambered galls as members of this single species. However, the support for three separate

lineages from nuclear and mitochondrial sequence data and the separation of two clades from the third by other recognised species suggest that there are in fact three cryptic species. Hartig's original specimens were collected in Central Europe, suggesting that the name *Andricus coriarius sensu stricto* rightly applies to the Main Clade, while the Lebanese and Iranian Clades represent new species.

The discovery of these taxa and their placement within *Andricus* has implications for the understanding of the evolution of gall phenotypes. Gall morphology is controlled by gallwasp genes, and its adaptive significance remains a subject of debate (Stone & Schönrogge 2003). *Andricus coriarius* (as previously understood) represented the only member of the *Andricus kollari* group of gallwasps to show two significant gall characteristics: the presence of many larval chambers (multilocularity) rather than a single chamber, and the presence of many surface spines (rather than galls that are spineless). Both of these traits are thought to play a role in gall defence against natural enemies (Stone & Schönrogge 2003). A reconstruction of character evolution through the genus *Andricus* (Stone & Cook 1998, using a sequence for *A. coriarius s.s.* from Hungary) inferred multilocularity and spininess both to represent derived states that evolved in *Andricus coriarius* after its divergence from an ancestor inducing single-chambered, spineless galls – character states present in the remaining members of the *Andricus kollari* group. The inference that two lineages inducing spiny, multi-chambered galls diverged basally to the remainder of the *Andricus kollari* group implies either (a) that each represents an independent evolution of both spininess and multilocularity or (b) that the spineless, single-chambered galls characteristic of the *Andricus* species comprising the sister group to *A. coriarius s.s.* are themselves derived. Resolution of these alternatives awaits formal reanalysis of gall phenotype evolution through *Andricus* as a whole. Independent convergent evolution of both spininess and multilocularity would be of particular interest, because it has occurred independently elsewhere in *Andricus* (Stone & Cook 1998) and in other gall-inducing insects (Stone & Schönrogge 2003). It has been proposed that natural selection has favoured the evolution of spines as a

defence against natural enemies attracted to the greater concentration of prey resources within a multilocular gall (Stone & Schönrogge 2003).

4.4.2 The geographic origin of *Andricus coriarius sensu stricto*

Although intraspecific phylogenetic analyses were not valid, the haplotype root probabilities derived from the statistical parsimony network support a geographic origin for *A. coriarius s.s.* to the east of Europe. This parallels the pattern seen in *Andricus quercustozae* (Rokas *et al.* 2003a). There were probably multiple glacial refugia in Asia Minor and the Caucasus, and the existence of multiple mountain range barriers to gene flow is thought to underlie much of the diversity of the region (Davis 1965-1985, 1971). A major faunistic and floristic divide in the region, the Anatolian Diagonal, runs from The Taurus Mountains in southeastern Turkey north-eastwards towards the Caucasus. Of the section *Cerris* oaks available for host-alternating *Andricus*, *Q. cerris* extends far to the west of this divide, while *Q. libani* is predominantly found to the south and east of it (Davis 1965-1985, 1971; Yaltirik 1982). One possible explanation for the eastern restriction of the Lebanese and Iranian Clades is that they exploit only *Q. libani* as a sexual generation host, and are unable to exploit *Q. cerris*. If true, such a limit to range expansion would mirror the inability of Iberian host-alternating *Andricus* to escape their glacial refuge by making an equivalent eastwards host shift from *Q. suber* to *Q. cerris* (Stone *et al.* 2001, Hayward & Stone 2006). In contrast, the geographic distribution of haplotypes of *A. coriarius s.s.* implies that basal members of this lineage were able to exploit both *Q. libani* (the only host available to Iranian populations) and *Q. cerris*, and to make a subsequent host shift to *Q. suber* when they reached Iberia.

4.4.3 Timescale and modes of dispersal

Andricus coriarius s.s., in common with *A. kollari* and *A. quercustozae*, diverged around 3.5 mya, before the Pleistocene glacial cycles (Table 4.4, Stone *et al.* 2001, Rokas *et al.* 2003a). The Iranian and Lebanese Clades are inferred to have more ancient origins in the early Pliocene and late Miocene, respectively. To an order of

magnitude (given the caveats associated with assuming a given rate of mitochondrial sequence evolution), the aspects of gall structure common to the three clades of *A. coriarius* have either been conserved in independently evolving lineages for almost 10 million years, or converged over a similar timescale. Vicariance patterns in oak floras in both sections *Cerris* and *Quercus s.s.* suggest that the three lineages currently known as *A. coriarius* diverged at the same time as their host oaks diversified into a characteristic Western Palaearctic flora (Manos & Stanford 2001). Although the Lebanese and Iranian Clades are both ancient lineages, each is inferred to have undergone a recent bottleneck, since diversity within these clades dates from the mid-Pleistocene (Table 4.4). A similar pattern is seen in Iberian populations of *Andricus kollari*, for which the MRCA is dated with the same methods at 0.4 mya (Hayward & Stone 2006).

Nested clade analysis of *A. coriarius s.s.* could not reject the null hypothesis of no geographic association of haplotypes for any clades at the first three nesting levels. Such a failure to detect significant associations can arise through insufficient sampling, panmixia, or lack of genetic variation (Templeton *et al.* 1995). Within the lower level clades, each of these explanations could be sufficient to account for the lack of significant associations. A benefit of the nested statistical design is that the statistical power of the NCPA is pooled at higher levels (Templeton 2004) so geographic associations can be detected in higher nesting clades. Colonisation of Spain (clade 4-3) can be explained by contiguous range expansion. The inference at the total cladogram level is past gradual range expansion with limited long-distance dispersal. The suggestions of contiguous range expansion are supported by the unimodal distribution of the pairwise mismatch plot of the main *A. coriarius* clade, which infers growth as a single demographic entity. Such a pattern has been associated with Western Palaearctic postglacial range expansion in other organisms, including woodmice (Michaux *et al.* 2003) and the Spined loach (Culling *et al.* 2006).

4.4.4 Elevation of cryptic lineages to species status

Diagnostic morphological differences have been identified between *A. coriarius s.s.* and the members of the Iranian and Lebanese clades (Melika *et al.* 2007). These differences support the inference from the molecular data presented in this chapter that the gallwasps identified as *A. coriarius* in this study represent three distinct species groups. The two species identified here as the Iranian and Lebanese clades have been named *Andricus coriariformis* (Melika, Challis & Stone 2007) and *Andricus libani* (Melika, Challis & Stone 2007), respectively.

Chapter 5

The phylogeographic clade trade: Tracing the impact of human-mediated dispersal on the colonisation of northern Europe by the oak gallwasp *Andricus kollari*†

Chapter 4 highlighted the importance of eastern diversity in the phylogeographic structure of an oak gallwasp that has been largely unaffected by human activities. This chapter uses allele frequency and DNA sequence data to distinguish between alternative scenarios (unassisted range expansion and long range introduction) for the colonisation of northern Europe by Andricus kollari, *an oak gallwasp whose distribution has been affected by human activity.*

5.1 Introduction

Human-assisted dispersal can be a powerful force in shaping the genetic structure of animal populations, superimposing new signatures on existing natural phylogeographic patterns. An understanding of the historic importance of organisms that have been widely traded, or are successful hitchhikers, is essential to unravelling their phylogeographic history. Human trade provides a dispersal mechanism most infamously exploited by the plague vector, *Xenopsylla cheopis* (the oriental rat flea), and phylogeographic methods have been used extensively in analyses of the origin(s) and dispersal routes of pests (e.g. Grapputo *et al.* 2005; Navia *et al.* 2005), disease organisms (e.g. Morgan *et al.* 2005; Lehrmann *et al.* 2006), disease vectors (e.g. Mousson *et al.* 2005; Fonseca *et al.* 2006) and domestic animals (e.g. Morii *et al.* 2002; Haag *et al.* 2004; Meadows *et al.* 2005).

Andricus kollari is one of nine members of the host-alternating *Andricus* clade of oak gallwasps (Section 1.4.6) whose distributions in the Western Palaearctic have been influenced by human activity (Stone & Sunnucks 1993; Stone *et al.* 2001, 2002; Csóka *et al.* 2005; Section 1.4.8). Ranges of these species have extended north of the latitudinal band in which both section *Cerris* and section *Quercus s.s.* oaks are

† A modified version of this chapter has accepted for publication as:
Stone GN, Challis RJ, Atkinson RJ, Csóka G, Hayward A, Mutun S, Preuss S, Rokas A, Sadeghi E & Schönrogge K (in press) The phylogeographic clade trade: Tracing the impact of human-mediated dispersal on the colonisation of Northern Europe by the oak gallwasp Andricus kollari, Molecular Ecology.

native (Sections 1.3.3) as human planting of *Q. cerris* has expanded the range of their sexual generation host into northern Europe (Section 1.3.5). Work on three of these range expanding species shows a gradual and continuous loss of population genetic diversity away from the native range (Stone & Sunnucks 1993; Csóka *et al.* 1998). Populations further along the invasion route show a subset of the allelic diversity found in populations closer to the native range. There is no evidence of the discontinuous distribution of genetic diversity predicted for large-scale introductions ahead of the invasion front, and no evidence that the English Channel presented any more of a barrier to dispersal than an equivalent distance over land. Previous work on *A. kollari* shows that the establishment of some populations in northwestern continental Europe can similarly be explained by range expansion by populations native to Italy and the Balkans, for which the introduced *Q. cerris* is also the natural sexual generation host. In contrast, Iberian populations have failed to expand beyond the northern limits of the natural distribution of cork oak, *Q. suber*, in the Gironde, southwestern France (Stone *et al.* 2001; Hayward & Stone 2006).

However, a compatible alternative explanation exists for *A. kollari* populations in Britain. *A. kollari* galls have been widely traded and, uniquely among the range-expanding gallwasps, parthenogenetic generation galls of this species were deliberately imported by sea to the southwest of Britain in the 1830s (Section 1.4.8). The *A. kollari* galls imported to Britain originated from a currently unknown region in the eastern Mediterranean and the import of *A. kollari* predated the arrival in Britain and northern continental Europe of naturally invading gallwasps by a century (Stone & Sunnucks 1993).

The first aim of this chapter is to use sequence and allele frequency data to reconstruct the native range phylogeography of *A. kollari*, extending previous work to include sampling in possible eastern sources of the galls imported to Britain. In particular, to determine whether data for *A. kollari* support an eastern origin within the Western Palaearctic, as inferred for other host-alternating *Andricus* species (Rokas *et al.* 2003a; Challis *et al.* 2007; Chapter 4).

The second aim of this chapter is to attempt to distinguish between three alternative hypotheses for the origin of current British *A. kollari* populations, that they are: (i) primarily derived from a continuation of the unassisted range expansion inferred for the invasion of continental Europe, (ii) primarily the descendants of the nineteenth century introductions from an area east of these sources, or (iii) a combination of these. These scenarios are expected to leave different phylogeographic signatures in *A. kollari*. In addition to the patterns in genetic diversity described above, unassisted range expansion is predicted to result from rapid population growth from small numbers of founders. Sequences sampled from the invaded range are expected to represent a subset of ancestral native range polymorphism, and any clades restricted to the invaded range are expected to have a very young most recent common ancestor (MRCA). In contrast, large-scale introduction is expected to import much of the genetic diversity present in the source population to a new location, creating a new area of high diversity in the introduced range that lacks a signature of rapid population growth from small numbers of founders. If eastern Mediterranean source populations are genetically distinct from those in the Balkans and Iberia, British populations are expected to be genetically distinct from those already examined in northern France (Stone *et al.* 2001; Hayward & Stone 2006). The MRCA of such introduced sequences is expected to long predate the 1830s introduction events. Establishment of a large population at the periphery of the invaded range also has the potential to result in dispersal from Britain into continental Europe, against the prevailing direction of unassisted range expansion.

5.2 Materials and Methods

5.2.1 Sample sites

The full longitudinal range of the natural distribution of *A. kollari* was sampled, from Portugal to Iran. No galls were found in Lebanon despite 14 man-days of sampling, and although recorded from Israel (Sternlicht 1968), this species is now extremely rare throughout the Levant. The only region where *A. kollari* could potentially be

found that has not been sampled extends from the extreme southeast of Turkey into the Kurdish autonomous region of Iraq, and is currently inaccessible.

Analyses of allozyme data include 2092 individuals from 69 sites. Of these, data for 1457 individuals from 46 sites in continental Europe (Spain, France, Germany, Holland, Italy, and Hungary) and Turkey were presented in Stone *et al.* (2001). Allele frequency summaries for these sites (23-49 and 51-69 in on-line supplementary material Appendix 1) are available at http://www.blackwellpublishing.com/products/journals/suppmat/mec/mec1211/mec1211sm.htm. This chapter presents data for a further 562 individuals from 20 sites in Britain, 59 from 2 sites in the Republic of Ireland, and 14 from 1 site in Belgium (see Appendices 3 and 4). The locations of all sites used in allozyme analyses are shown in Figure 5.1.

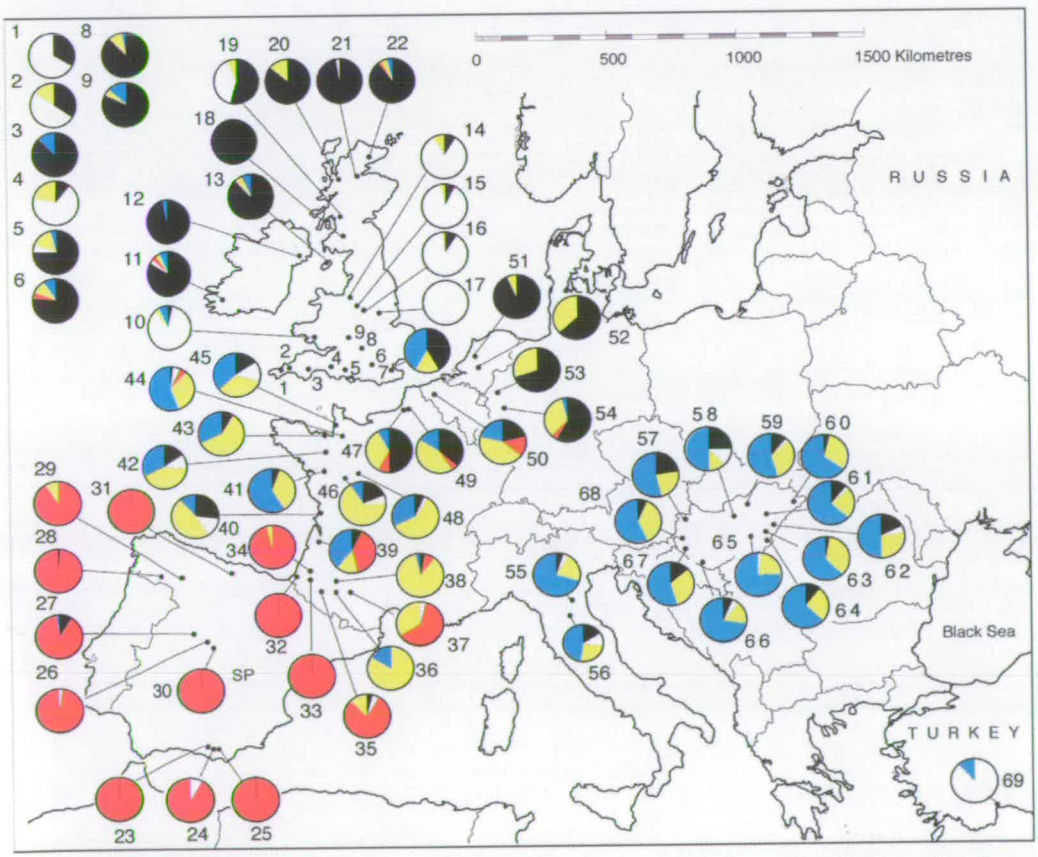


Figure 5.1 The proportions of individuals in each population allocated to the 5 populations identified by Structure (1=red, 2=blue, 3=white, 4=yellow, 5=black). Site numbers correspond to those in Table 5.1.

Sequences for fragments of the mitochondrial cytochrome b (*cytb*) gene and the D2 expansion region of the nuclear 28S gene (see below) were obtained for 160 individuals from 53 sites and 56 individuals from 27 sites, respectively (see Appendix 5). Analyses of *cytb* data include sequences for 54 individuals published previously in Stone *et al.* (2001; Genbank accession numbers AF242739-AF242762 and AF242764-AF242766) and Hayward & Stone (2006; DQ925335-DQ925361). The final *cytb* dataset was distributed across sampled countries as follows (with numbers of individuals sequenced in parentheses): Spain (20), Portugal (1), France (52), UK (19), Ireland (3), Holland (6), Germany (1), Italy (16), Hungary (15), Turkey (22), Iran (3).

5.2.2 Allozyme screening

Individual wasps were homogenised in allozyme extraction buffer (Peakall & Beattie 1991) and scored for allelic variants at 13 polymorphic loci using cellulose acetate electrophoresis (Zip-zone, Helena Laboratories) and substrate-staining protocols described by Richardson *et al.* (1986) and as used in Stone & Sunnucks (1993) and Stone *et al.* (2001). AK, αGPD1, αGPD2 and PEP-b were run on 40mM sodium phosphate pH6.3 (Stone & Sunnucks 1993), GOT-s, GOT-m, GPI, MDH-s, MDH-m, ME, 6PGD were run on 0.1M Tris-EDTA-maleate-$MgCl_2$ pH 7.6 (Richardson *et al.* 1986, buffer F) and HK and PGM were run on 25mM Tris-Glycine pH 8.5 (Richardson *et al.* 1986, buffer I).

5.2.3 Analyses of allele frequency data

Genotypic data were tested for deviations from Hardy-Weinberg equilibrium using the approach of Guo & Thompson (1992) incorporated in Arlequin 3.0 (Section 2.12.1.1) with default settings, and for linkage equilibrium using the permutation procedure in Genetix 4.0 (Section 2.12.6.1). Significance levels in both cases were adjusted for multiple tests using a Bonferroni correction (corrected threshold p value = $1-(1-\alpha)^{1/k}$ where k is the number of tests and α is the desired threshold value of 5%, as in Stone *et al.* (2001).

Relationships between populations were inferred using neighbour joining analysis (NJBP, Jean-Marie Cornuet, INRA Laboratoire de Modélisation et Biologie Évolutive, Montpellier) of Cavalli-Sforza's chord distance (Cavalli-Sforza & Edwards 1967) to allow comparison with previous work on this system (Stone *et al.* 2001). Bootstrapping was carried out over both populations and loci, with 1000 bootstrap replicates. Pairwise F_{st} values between populations were calculated in Arlequin 3.0 (Section 2.12.1) and their significance assessed using the permutation procedures in this programme.

The number of discrete populations (genotype pools) in *A. kollari* was determined using the programme Structure (Section 2.12.13). Simulations for K (number of populations) = 1 to 6 were run for 1 x 10^6 generations with a burn-in of 1 x 10^5 generations, and convergence in the estimated parameter values was checked over two independent runs at each K. Results with and without admixture were indistinguishable (as indicated by the small value of the Dirichlet parameter) so only the results without admixture are reported.

5.2.4 DNA extraction, amplification and sequencing

DNA was extracted using the DNeasy Tissue kit (Section 2.11.2). 433 bp fragments of the mitochondrial cytochrome *b* (*cytb*) gene and 489 bp fragments of the nuclear ribosomal 28S gene D2 region (28SD2) were amplified (Sections 2.11.3.1 and 2.11.3.3) and sequenced in both directions (Section 2.11.4). All new sequences were deposited in Genbank (accession numbers EF030046-EF030047, 28SD2, and EF031335-EF031457, *cytb*).

5.2.5 Phylogenetic analysis

Because only two 28S D2 haplotypes were found over the full range of *A. kollari*, formal phylogenetic analysis was only carried out on the *cytb* sequences. The validity of applying tree-based analyses to this dataset was tested using the method of Huson & Bryant (2006) (Section 2.7.3) in SplitsTree 4.4 (Section 2.12.12). Distances were

estimated under the HKY+I+G model of evolution using PAUP* 4b10 (Section 2.12.11) and model parameters estimated using MrModeltest (Section 2.12.7). Base frequencies: A = 0.3743, C = 0.1167, G = 0.0913, T = 0.4177; ti/tv ratio = 5.7454, I = 0.5302; Gamma distribution shape parameter = 0.5510). The hypothesis of tree-like evolution was rejected for the *cytb* data and relationships between haplotypes are therefore represented using the SplitsTree network. Before inferring relationships and demographic histories with these data, the absence of any signature of selection was confirmed using the McDonald-Kreitman test (Section 2.12.5).

5.2.6 Inference of population demographic history

Population demographic history was assessed using pairwise mismatch distributions (Section 2.12.1.2). For lineages showing significantly unimodal distributions, Arlequin was used to estimate relative population sizes ($2\mu N$) before (θ_0) and after (θ_1) population growth, the relative time since the onset of population expansion ($\tau=2\mu t$), and Fu's F_S statistic (Section 2.12.5.1).

5.3 Results

5.3.1 Analyses of allozyme allele frequency data

5.3.1.1 Hardy Weinberg and Linkage equilibria

Over all 69 populations, all 13 of the allozyme loci were polymorphic, yielding a total of 64 alleles. Allele frequencies for sites in Britain and Ireland (sites 1-22) and Belgium (site 50) are presented in Appendix 4. Over all 69 populations, there were small numbers of significant departures from Hardy-Weinberg equilibrium (HWE: 14 of 484 tests) and linkage equilibrium (74 of 5850 pairwise tests). As in prior analyses of these loci in this species (Stone *et al.* 2001), departures from these equilibria were consistent neither across loci within a given population, nor across populations for a given locus and both equilibria have been assumed in the structure analysis.

5.3.1.2 Geographic patterns in allele frequencies

Variation in allele frequencies resolves *A. kollari* populations into 4 groups (Fig.2). The native distribution of *A. kollari* resolves into 2 groups: (1) a strongly supported group containing all populations from the native range of *Q. suber* in Spain and southwest France; and (2) a grouping comprising southern central Europe (Italy and Hungary in the native range of *Q. cerris*) and most populations in northwestern France. United Kingdom populations are divided between two further groupings: (3) a group comprising 8 UK populations and the Turkish population of Antalya; and (4) 12 UK populations with neighbouring regions of northern coastal Europe (Republic of Ireland, Belgium, Holland, Germany and the two French sites of Crécy [49] and Nouvion [47] on the Channel coast).

While the two groups of UK populations are not geographically separated (both groupings comprise sites distributed from southern England to Scotland), they are genetically very divergent (Table 5.1a). While pairwise F_{st} between sites within each group is low and sometimes not significantly different from 0, values between sites in the two groups have a mean of 0.36. Pairwise F_{st} values between regions (Table 5.1b) show that the UK populations grouping with Turkey in Figure 5.2 are relatively divergent from all regional groupings of populations, including Turkey. Those UK populations clustering with neighbouring regions of northern coastal Europe in Figure 5.2 are very divergent from the Turkish population, but relatively less divergent from those in Hungary and Italy.

Table 5.1 (a) Pairwise F_{st} between sites in the two groups of UK populations, shown for four representative sites in each group. UK Sites whose allele frequencies cluster them with Turkey in Fig. 2 and whose individuals are predominantly in Structure population 3 are Chatsworth, Knutsford, Rufford Park and Swansea. UK sites affiliated to populations in northern coastal Europe and and whose individuals are predominantly in Structure population 5 are Beauly, Puttenham Common, Oxford and Skares. All values in bold type are significantly different from zero at $p<0.05$ using the permutation procedure in Arlequin. (b) Mean pairwise F_{st} ± standard error (sample size) for the same two sets of UK populations and representative populations in continental northern coastal Europe (Torhout, Utrecht, Amsterdam, Köln), northwestern France (Bercé, Coutances, Crécy, St. Malo and Rennes), Hungary (Fehérgyarmat, Hortobágy, Mátrafüred, Nagykanizsa and Zalaegerszeg) and Turkey (Antalya). All values between sets of populations are significantly non-zero at $p<0.01$ using the permutation procedure in Arlequin.

(a)

	STRUCTURE population 3			STRUCTURE population 5			
Site name	**Chat.**	**Knut.**	**Ruff.**	**Swan.**	**Beau.**	**Putt.**	**Oxfo.**
Knutsford	0.014						
Rufford	**0.020**	**0.055**					
Swansea	**0.048**	**0.064**	**0.067**				
Beauly	**0.379**	**0.361**	**0.449**	**0.329**			
Puttenham	**0.297**	**0.276**	**0.364**	**0.237**	0.025		
Oxford	**0.333**	**0.311**	**0.406**	**0.277**	**0.030**	0.003	
Skares	**0.369**	**0.331**	**0.446**	**0.328**	0.018	**0.025**	0.037

(b)

Region	**UK sites in population 3**	**UK sites in population 5**
UK Sites in population 3	0.045±0.008 (6)	
UK Sites in population 5	0.360±0.015 (12)	0.023±0.004 (6)
Continental northern coastal Europe	0.261±0.019 (16)	0.105±0.020 (16)
Northwestern France	0.181±0.011 (20)	0.115±0.010 (20)
Hungary	0.200±0.009 (20)	0.114±0.008 (20)
Turkey	0.169±0.016 (4)	0.375±0.020 (4)

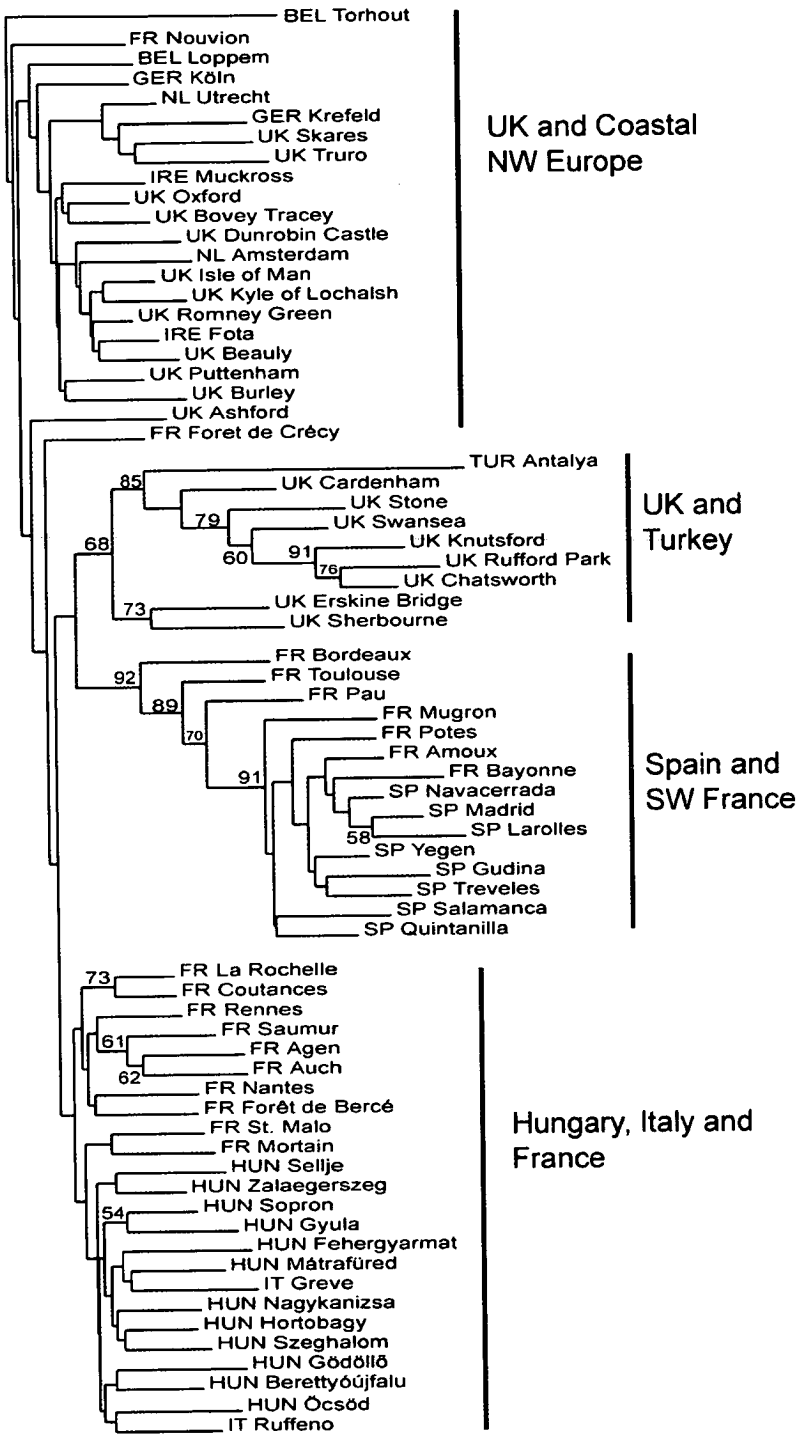


Figure 5.2 An arbitrarily rooted phylogram of relationships based on allele frequency data among *A. kollari* populations. Relationships are generated by neighbour joining of Cavalli-Sforza and Edward's chord distance. Numbers at nodes are bootstraps, over individuals and populations, expressed as a percentage of 1000 replicates. Nodes without bootstraps are supported by less than 50% of replicates.

Both groups of UK populations are significantly differentiated from populations representing possible sources for unassisted range expansion from Iberia or central Europe. The main genetic divide between the UK and continental Europe is at the Channel coast, and is apparent in the distribution of regionally private alleles. Sites at the Channel coast of France have 4 alleles that are shared with Italy and Hungary (αGPD2 allele 1, GOT-m allele 5, PEP-b allele 8, PGM allele 5) but absent from Britain and northern coastal Europe. This northwards loss of alleles is potentially compatible with founder effects during unassisted range expansion. However, UK and Irish populations also possess six alleles that are absent from northwestern France (GOT-s allele 1, αGPD1 allele 1, HK allele 3, AK alleles 1, 2, 4). Significantly, AK allele 1 (found at 9 sites in the UK and Ireland, reaching a frequency of 40% at Knutsford) and AK allele 4 (only found as 6 copies in Cardinham, UK) are otherwise absent from the entire sampled range of *A. kollari*.

5.3.1.3 Analyses of individual multilocus genotypes

Structure analysis strongly supports the existence of 5 populations in *A. kollari* (Figure 5.1). The posterior probability for k=5 is ~1 and for all other values of k is ~0. Native range *A. kollari* are allocated predominantly to populations 1 (Iberia and southwestern France), 2 (Italy and Hungary) and 3 (Turkey). Population 4 contains individuals from sites throughout the native and introduced range of *Quercus cerris*. Population 5 primarily contains individuals from the UK and neighbouring regions of northern coastal Europe (Ireland, Belgium, the Netherlands and Germany), but is also represented in northern France, Italy and Hungary (Figure 5.1).

This approach shows the strong genetic similarity between populations in northwestern France and the native range of *Q. cerris* in Italy and Hungary. Very few individuals sampled outside the native range of *Q. suber* are allocated to the Iberian population: of 359 individuals sampled in northern France, Belgium, Holland and northern Germany, only 8 (2.2%) were allocated to the Iberian population 1, and these had low allocation probabilities (mean 0.51, range 0.37 - 0.71). Dispersal from

the native and introduced ranges of *Q. cerris* into the Iberian Peninsula is also inferred to be very rare: Of 284 individuals sampled in Spain, only 10 were not allocated to the Iberian population 1 (4,3,2 and 1 individuals were allocated to populations 3,2,4 and 5 respectively, with probabilities ranging from 0.30 to 0.80).

British sites fall into three groups on the basis of their Structure allocation (Figure 5.1). One set (8 sites) contains sites for which a majority of individuals are allocated to the Turkish population 3. A second (13 sites) contains sites for which a majority of individuals are allocated to the fifth population. This division corresponds to the contrast between two groups of UK sites discussed above. The third group consists of Ashford in southeastern England, which is exceptional in having a high proportion of individuals allocated to population 2, thus resembling sites in northwestern France. Across Britain, only small minorities of individuals were allocated to the Iberian population 1 (4 individuals, 0.7%), and populations 2 (23 individuals, 4%) and 4 (37 individuals, 6.6%) common in the European native range of *Q. cerris*.

Sites in neighbouring regions of northern coastal Europe showed strong similarity to those British populations dominated by the fifth population, but contained very few individuals allocated to Turkish population 3 (Figure 5.1).

5.3.2 Analyses of sequence data

5.3.2.1 Phylogeographic patterns in cytochrome *b* and 28S D2

The *cytb* sequencing added 62 new haplotypes to those recorded by Stone *et al.* (2001) and Hayward & Stone (2006), giving a total of 108 *A. kollari* haplotypes. The haplotypes recorded at each site are presented in Appendix 5. The entire dataset of 108 discrete haplotypes has 94 polymorphic sites, 49 of which are parsimony informative.

The *cytb* network (Figure 5.3) shows 4 major lineages diverging from an unresolved polytomy.

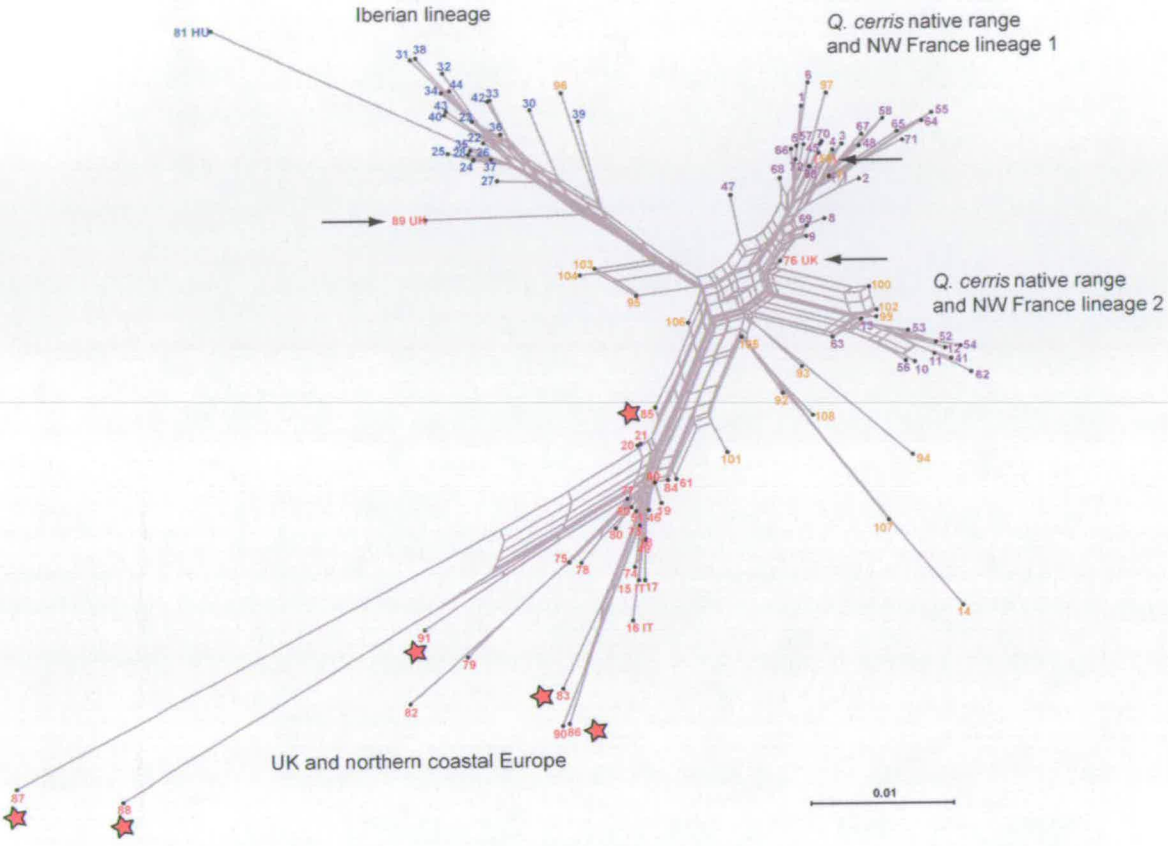


Figure 5.3 A Neighbor-Net network (Bryant & Moulton 2004) of *cytb* haplotypes. The network was constructed using an equal angle splits transformation (Dress & Huson 2004) of distances under the HKY model of evolution in SplitsTree 4.4 (Huson & Bryant 2006). Geographic regions represented in each lineage are colour coded as follows: native range of *Q. suber* (blue), native range of *Q. cerris* in Europe and the introduced range of *Q. cerris* in France (purple), Turkey and Iran (orange) and the UK and northern coastal Europe (red). Haplotypes marked with a star indicate individuals whose multilocus allozyme genotypes were allocated to Structure population 3. Haplotypes marked with an arrow indicate individuals sampled in the UK and northern coastal Europe that fall outside the main lineage for this region.

All haplotypes from the native range of *Q. suber* form a single discrete lineage, which also contains three haplotypes sampled outside Iberia – one from Hungary (haplotype 81 from Gyula), one from the UK (haplotype 89 from Swansea) and one from Turkey (haplotype 96 from Bolu). Haplotypes from the native range of *Q. cerris* lie predominantly in two major lineages, each encompassing Central

Europe and Turkey. The native range haplotypes not in these 3 major groups are from Turkey and Iran, and diverge independently from the central polytomy in the network. Across the native range, nucleotide diversity is greatest for Turkey (0.0200), intermediate in Italy (0.0151) and Hungary (0.0136), and lowest in Spain (0.0068).

The introduced range of *Q. cerris* in France contains haplotypes from both of the lineages representing the native range of this oak, implying independent colonisation of northwestern Europe by members of each lineage. Contributions from both *Q. suber*- and *Q. cerris*-associated lineages mean that France contains the highest nucleotide diversity (0.0224) found in any single region.

With 3 exceptions, haplotypes from sites in the UK and northern coastal Europe (Ireland, Belgium, Holland and northern Germany) form a discrete lineage with high nucleotide diversity (0.0170). This lineage also includes 2 haplotypes sampled from sites in northern France (haplotypes 15 from Crécy and 82 from Nantes). Though the UK sequences form a discrete group relative to those sampled elsewhere, there is evidence that they comprise two genetically divergent populations. Individuals whose allozyme genotypes (comprising 11 nuclear and 2 mitochondrial loci) are allocated to Structure population 3 are significantly over-represented among the most divergent haplotypes (marked with stars in Figure 5.3), while individuals allocated to Structure population 5 are significantly over-represented among the less divergent sequences in this lineage (G-test for association using Williams' correction, G_{adj}=8.00 on 1 degree of freedom, $p<0.005$; Sokal & Rohlf 1981).

Of the three UK and northern coastal Europe sequences lying outside the UK+northern coastal Europe lineage (marked with arrows in Figure 5.3), one falls in the Iberian lineage (haplotype 89 from Swansea, UK), and two are associated with *Quercus cerris* native range lineage 1 (haplotypes 59 from Utrecht, Holland, and 76 from Cardinham, UK).

The D2 region of 28S revealed only 2 haplotypes, one present throughout the native range of *Q. suber* in Spain and southwestern France, and one shared by all other individuals. The individuals from Gyula and Bolu whose cytochrome b haplotypes lie in the Iberian lineage both had non-Iberian D2 haplotypes while the individual from Swansea had one copy of each of the D2 haplotypes. The two of these 3 individuals also genotyped for allozymes had non-Iberian multilocus genotypes: the Swansea individual was allocated to population 2, and the Gyula individual to population 5, rather than Iberian population 1.

5.3.2.2 Population demographic history

Of the lineages shown in Figure 5.3, two showed a significantly unimodal pairwise mismatch distribution indicating rapid population expansion (see Fig. 5 for Iberia): for *Q. cerris* native range lineage 2 SDD=0.0455, p=0.05; for Iberia, SSD=0.0916, p<0.001. Assuming that μ is equivalent in the two lineages, the population expansion in the native range of *Q. cerris* was more dramatic (θ_0=0.025, θ_1=3.63, for the Iberian lineage θ_0=0.388, θ_1=1.159) and twice as long ago (τ=2.559, for the Iberian lineage τ=1.248). Significant population growth for both of these lineages was also inferred by Fu's Fs test: F_S = -8.143, $p < 0.001$ for *Q. cerris* native range lineage 2, and F_S = -10.674, $p < 0.001$ for the Iberian lineage.

The UK and northern coastal Europe lineage has a multipeaked mismatch distribution more characteristic of a large and stable population (Figure 5.4). No signature of population expansion was detected (F_S = -2.278, p = 0.169). The same was true for this clade when the more divergent sequences corresponding to Structure population 3 (identified in Figure 5.3) were excluded (F_S = -0.555, p= 0.35).

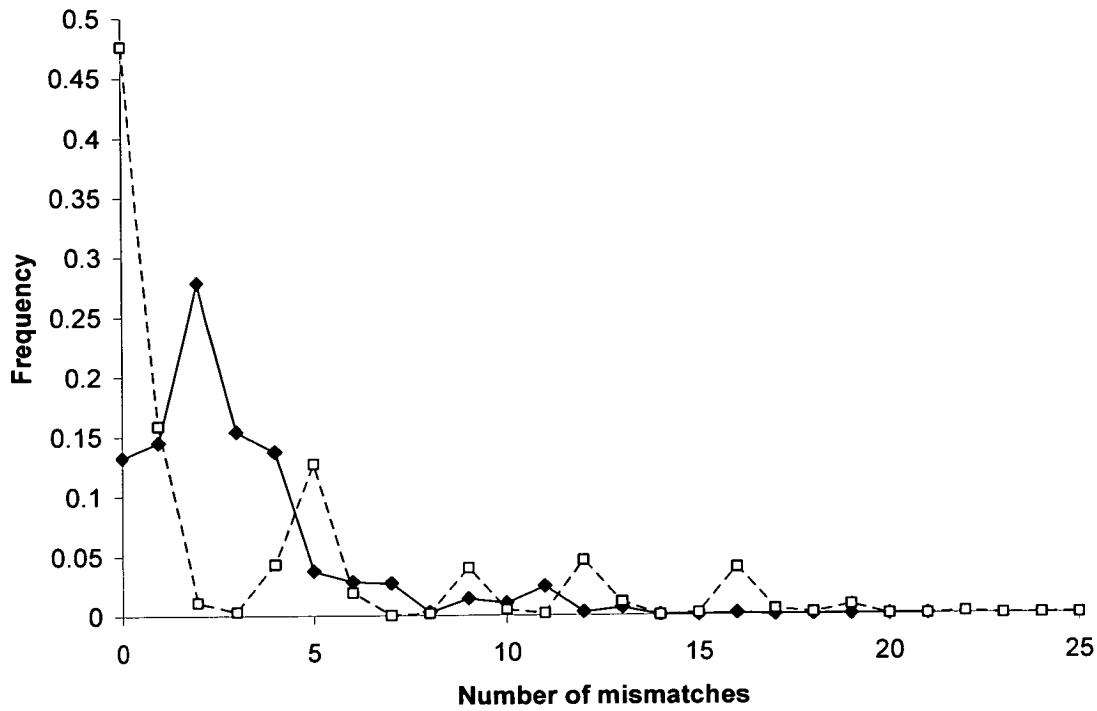


Figure 5.4 Pairwise mismatch plots for *cytb* sequences in the Iberian (filled diamonds) and UK and northern coastal Europe (open squares) groups in Figure 5.3.

5.4 Discussion

5.4.1 Native range phylogeography of *Andricus kollari*

Data for two other host-alternating *Andricus* species (*A. coriarius* and *A. quercustozae*) support the hypothesis that European populations are derived from lineages originating in Turkey or Iran, leading to the hypothesis of an eastern cradle for the host-alternating clade of *Andricus* gallwasps (Rokas *et al.* 2003a; Challis *et al.* 2007; Chapter 4). The lack of resolution of the relationships between lineages in the *A. kollari* network mean that a similar 'out of the east' signature in *A. kollari* cannot be confirmed or rejected. However, the high nucleotide diversity in Turkey and Iran, and the star-like distribution of eastern sequences around the central polytomy in the haplotype network are both compatible with an eastern origin. The signature of population growth in one of the lineages containing the European native range of *Q. cerris* matches that demonstrated for *Andricus coriarius* (Challis *et al.*

2007; Chapter 4). In *A. coriarius,* this signature was inferred to be the result of westwards range expansion 1.6 million years ago. The signature in *A. kollari* is compatible with a similar event.

All of the approaches used here confirm earlier work (Stone *et al.* 2001; Hayward & Stone 2006) showing a major divide between Iberian populations galling cork oak, *Q. suber* and those populations galling turkey oak, *Q. cerris*. In *A. coriarius* and *A. quercustozae* Iberian haplotypes (and hence a lifecycle involving *Q. suber*) comprise a monophyletic clade derived from more eastern lineages associated with *Q. cerris* (Rokas *et al.* 2003a; Challis *et al.* 2007; Chapter 4). This relationship is compatible with the network obtained for *Andricus kollari*, but is not strongly supported by it. The low nucleotide diversity and signature of recent population growth detected in Iberian haplotypes could date from colonisation of *Q. suber* from *Q. cerris* by a small number of founders, followed by population expansion over a newly available resource. Alternatively, it could indicate recovery from a subsequent population bottleneck.

5.4.2 The origins of British *Andricus kollari* and cross-channel gene flow

Almost all *cytb* sequences for UK samples belong to a single lineage that includes samples from neighbouring coastal Europe but excludes all other sampled regions. This lineage has a pairwise mismatch distribution characteristic of a large and stable population, a signature that can only be explained by large-scale introduction from somewhere in the native range of this species. Extensive differentiation between UK sites and those in France, Iberia and southern central Europe is also supported by the allozyme data. This pattern contrasts strongly with three other host-alternating *Andricus* gallwasps (*A. corruptrix, A. lignicolus* and *A. quercuscalicis*) that colonised Britain without direct human assistance by dispersal from France (Stone & Sunnucks 1993; Csóka *et al.* 1998). The allozyme data for *A. kollari* suggest some affinity between a subset of UK populations and the single Turkish site sampled for these markers, but pairwise F_{st} between these populations is still substantial. All of these lines of evidence suggest that UK populations are primarily derived from one or

more large and genetically diverse introductions from unsampled parts of the eastern native range of *A. kollari*.

In addition to this primary route, haplotype sequences and Structure analysis provide evidence of rare gene flow into Britain across the English Channel from both *Q. suber* (haplotype 89, Swansea) and *Q. cerris*-associated lineages (haplotype 76, Cardinham) in continental Europe. These haplotypes probably reached Britain by continuation of the unassisted range expansion process that brought related haplotypes from Italy and Hungary into northwestern France. Direct dispersal from Spain is unlikely for the Swansea sample because although its mitochondrial haplotype is Iberian, one of its alleles for 28S D2 and its allozyme genotype are strongly diagnostic of *Q. cerris* regions outside Iberia (the allocation probability of this individual to the central European population 2 by Structure is 0.966). This suggests that this mitochondrial haplotype escaped Iberia by hybridisation near where the distributions of *Q. suber* and planted *Q. cerris* meet in southern France, followed by multiple generations of backcrossing to central European wasps to result in a non-Iberian nuclear genotype. The existence of a UK site showing substantial similarity in population allele frequencies to sites in France (Ashford in southeastern England, Figs.2, 3) suggests that successful colonisation from France does occur, as inferred for three other host-alternating gallwasps (Stone & Sunnucks 1993; Csóka *et al.* 1998). Cytochrome *b* sequence and allele frequency data show that from Britain *A. kollari* spread westwards into Ireland and eastwards into Belgium, Holland and northern Germany. Dispersal from Britain southwards into France appears to have been much more restricted, and only 2 of 30 haplotypes sampled from northern France fell within the UK and northern coastal Europe clade.

Genetic discontinuity across the English Channel, and between France and northern coastal Europe is thus maintained despite bidirectional gene flow. Genetic discontinuities are not present in the same area in other closely related and ecologically similar invading gallwasps (Stone & Sunnucks 1993; Csóka *et al.* 1998), suggesting that the pattern in *A. kollari* is unlikely to be a consequence either of its

biology or the distribution of its oak hosts. The pattern is consistent with predictions of models in which long-range dispersal ahead of an invasion front establishes spatial patterns in genetic diversity that are resistant to perturbation by subsequently arriving genotypes (Ibrahim *et al.* 1996; Lee & Hastings 2006). Even with rare long-range dispersal events, the patterns established in this way can persist for hundreds of generations. In the case of *A. kollari*, the number of introduced individuals was very large, and also (taking other *Andricus* invasions as a guide) occurred around 100 years before the unassisted arrival of central European genotypes. Both of these aspects of the introduction of *A. kollari* should extend the persistence of the resulting spatial patterns in genetic diversity.

5.4.3 How many sources?

The membership of a single lineage by all but 2 of the UK-sampled haplotypes for *cytb* suggests a single origin for most UK individuals. However, the correlation between nuclear allozyme genotypes and mitochondrial haplotypes within this lineage suggests that British *A. kollari* are derived from two related sources between which there has been limited gene flow in the past. Mean pairwise F_{st} values between the two main populations within the UK and northern coastal Europe clade (0.36) are greater than those between Hungary and Spain (0.32, Stone *et al.* 2001), indicating differentiation at the level of discrete glacial refugia. There are two possible unsampled or little-sampled regions that could harbour such diversity, and yet be related genetically. One is south of the Taurus Mountains in Turkey and Syria (Fig. 1). The Taurus Mountains are part of a major faunistic and floristic divide termed the Anatolian Diagonal (Davis 1965-1985, 1971; Çiplak 2003, 2004), which is also associated with major genetic divides in *Andricus* gallwasps (Rokas *et al.* 2003a). Sampling from this divide eastwards for *A. kollari* is limited to 10 haplotypes from sites near Kayseri, Bitlis and Mus (haplotypes 99-108 in Fig. 4). None of these haplotypes fall within the UK and northern coastal Europe lineage, but haplotype 104 (from Gevas, near Bitlis) is the most similar in sequence to this lineage of any native range samples. A second possible source is Lebanon and Israel, whose highlands support endemic floras (Zohary 1966; Shmida 1984) and which were probably

isolated enough during Pleistocene ice ages to constitute discrete refugia for oaks, and hence associated gallwasps. *A. kollari* could not be found in this region, but the possibility remains that they were present in the nineteenth century. Oak woodlands are under severe threat through degradation for charcoal production and by livestock throughout this region (Pons & Quézel 1985), and it is possible that one or both of the related sources for British *A. kollari* may be extinct in their region of origin.

5.4.4 How many introductions?

Unless gallwasps from the two inferred sources are able to mate assortatively after their introduction to Britain, associations between mitochondrial haplotypes and multilocus nuclear genotypes would be expected to have broken down rapidly through interbreeding. The fact that this has not occurred suggests that initial colonisation of Britain was by discrete pulses of individuals, each dominated by one of the two sources. Variation in the source responsible for colonising specific regions of the UK early in the invasion process could have established spatial patterning in the genetic make up of populations that has persisted to the present day. Such patchiness in the source of founding individuals is unlikely to have occurred were the two sources mixed in an individual shipment, and so spread together after their escape. It would be more likely if the source of gallwasps released in Britain varied between years. The trade in galls in the eastern Mediterranean was conducted through major trading centres such as Aleppo (Halab) in Syria that served a wide hinterland in Turkey, Syria, Iraq and Lebanon (Niebuhr 1776-1780). It is likely that the source of shipments exported through such centres varied from year to year.

Chapter 6

Evidence for genetic introgression in the phylogeography of the *Andricus quercuscalicis* clade

Chapters 4 and 5 concerned species that could be classified into discrete monophyletic taxa. This chapter investigates the phylogeography of molecularly and morphologically defined taxa in a clade of oak gallwasps in which the currently recognised species are non-monophyletic for mitochondrial DNA.

6.1 Introduction

Phylogenetic studies of the Western Palaearctic oak gallwasp fauna have divided the host-alternating *Andricus* species (Section 1.4.6) into four major clades: *A. hartigi, A. kollari, A. mayri* and *A. quercuscalicis* (Stone & Cook 1998; Cook *et al.* 2002; Rokas *et al.* 2003b). Species in each of these clades have been inferred to be non-monophyletic for the mitochondrial (mtDNA) cytochrome *b* (*cytb*) gene (Rokas *et al.* 2003b). However, the extent of discordance between *cytb* phylogeny and morphological taxonomy is greatest for the *A. quercuscalicis* clade.

Three out of four species in the *A. quercuscalicis* clade for which sequence data were available for more than one specimen (*A. caputmedusae*, *A. dentimitratus* and *A. mitratus*) were inferred to be non-monophyletic on the basis of mtDNA (Rokas *et al.* 2003b). Intrarefugial introgression (which predicts that mtDNA haplotypes will cluster by geography while nuclear markers may cluster by species species) was proposed as the most probable explanation for the lack of species monophyly. Since the adult gallwasps of such closely related species can be difficult to distinguish (and initial identification is typically made on the basis of gall morphology), an alternative explanation may be retained ancestral polymorphism in gall morphology. Under this alternative hypothesis, both nuclear and mtDNA markers are expected to cluster by geography. However, although this study included most members of the *A. quercuscalicis* clade, each species was represented

by only between one and three individuals, resulting in omission of intraspecific diversity. One of the aims of this chapter is to investigate the extent of possible introgression or retained ancestral polymorphism in gall morphology using more extensive geographic sampling.

Gall morphology is largely controlled by gallwasp genes and has adaptive significance (Section 1.4.4). The galls of the *A. quercuscalicis* clade (Figure 6.1) typically contain a single larval chamber, partially surrounded by an air space, and most have a sticky surface coating (Stone & Cook 1998). However, some species show regional variation. An aim of this chapter is to use more extensive sampling to investigate whether the regionally distinct gall morphologies of *A. insana* and *A. quercustozae* (Figures 6.1G and K, respectively; Section 1.4.4) are induced by members of separate, monophyletic lineages.

The natural distributions of the members of the *A. quercuscalicis* clade are restricted to the latitudinal band in which the native ranges of both section *Cerris* and section *Quercus s.s.* oaks overlap (Section 1.3.3). Within this band, longitudinal divisions in the distributions of section *Cerris* oak species (Section 1.3.4) may have promoted regional differentiation. Regional forms of oak gallwasp species outside of the *A. quercuscalicis* clade are increasingly being elevated to distinct species status. Differences in the phenology of *Q. suber* and *Q. cerris*, for example, have been demonstrated to present a reproductive barrier to sexual generations of *A. kollari* (Stone *et al.* 2001), which was followed by formal restoration of the Iberian lineage, *A. hispanicus*, to sibling species status (Pujade-Villar *et al.* 2003). Patterns in genetic diversity have also been used in *A. coriarius* to suggest the presence of two cryptic eastern lineages (Challis *et al.* 2007; Melika *et al.* 2007; Chapter 4). Further elevation of regional forms to species status has resulted from independent development of novel morphological characters to differentiate the adult gallwasps (Melika *et al.* 2007). A further aim of this chapter is to determine whether mtDNA sequence data provide evidence for the elevation of regional forms within the *A. quercuscalicis* clade to species status.

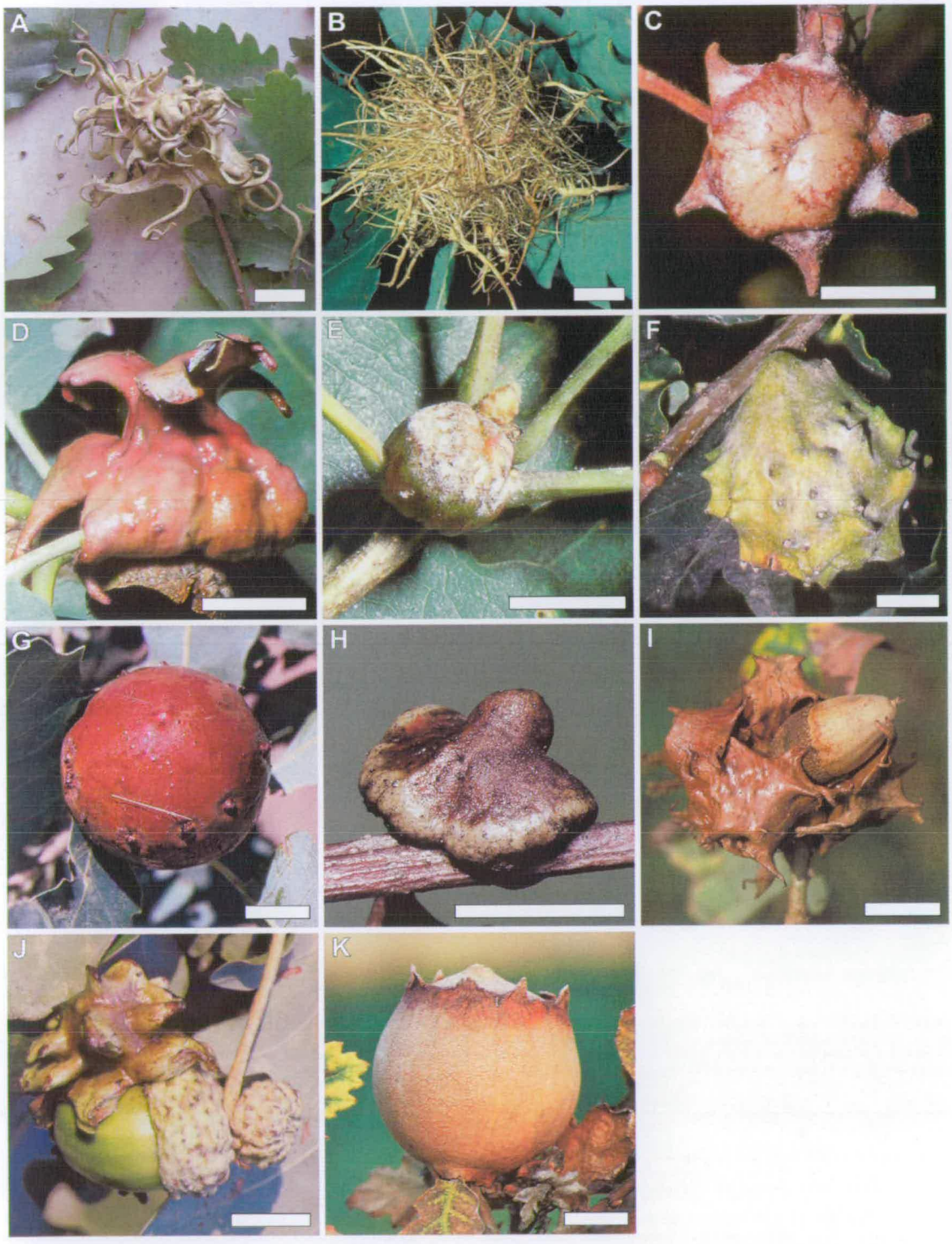


Figure 6.1 Parthenogenetic generation gall morphologies of the *Andricus quercuscalicis* clade: (A) *A. assarehi*, (B) *A. caputmedusae*, (C) *A. coronatus*, (D) *A. dentimitratus*, (E) *A. glutinosus*, (F) *A. hungaricus*, (G) *A. insana*, (H) *A. mitratus*, (I) *A. pictus*, (J) *A. quercuscalicis*, and (K) *A. quercustozae*. All scale bars are 10 mm.

This chapter uses extensive geographic sampling of four members of the *A. quercuscalicis* clade (*A. caputmedusae*, *A. dentimitratus*, *A. quercuscalicis* and *A. quercustozae*), together with published sequence data for a further five species (*A. askewi*, *A. coronatus*, *A. glutinosus*, *A. mitratus* and *A. pictus*) to attempt to answer the following specific questions:

1. Do sequences within the *A. quercuscalicis* clade predominantly cluster by species (according to current definitions) or by geography?
2. Are geographic divisions within the *A. quercuscalicis* clade consistent with those previously reported for other oak gallwasp species?
3. Is the distinction between the two gall morphotypes of *Andricus quercustozae* reflected in the phylogeny?
4. Is intrarefugial introgression or retained ancestral polymorphism in gall morphology more likely to explain the polyphyly of individual species (as currently defined)?
5. Under an intrarefugial introgression hypothesis, how frequent have introgression events been?
6. Can molecular taxa be resolved using DNA barcoding in a group that is subject to genetic introgression?

6.2 Methods

6.2.1 Sample collection

Sexual and parthenogenetic generation galls of four members of the *A. quercuscalicis* clade (*A. caputmedusae*, *A. dentimitratus*, *A. quercuscalicis* and *A. quercustozae*) were collected from locations across the species' ranges from Iberia to Iran (Figure 6.2). As far as possible, sample sizes were balanced across species ranges. Regions without sampling for a given species reflect extreme rarity of absence of that species from that region.

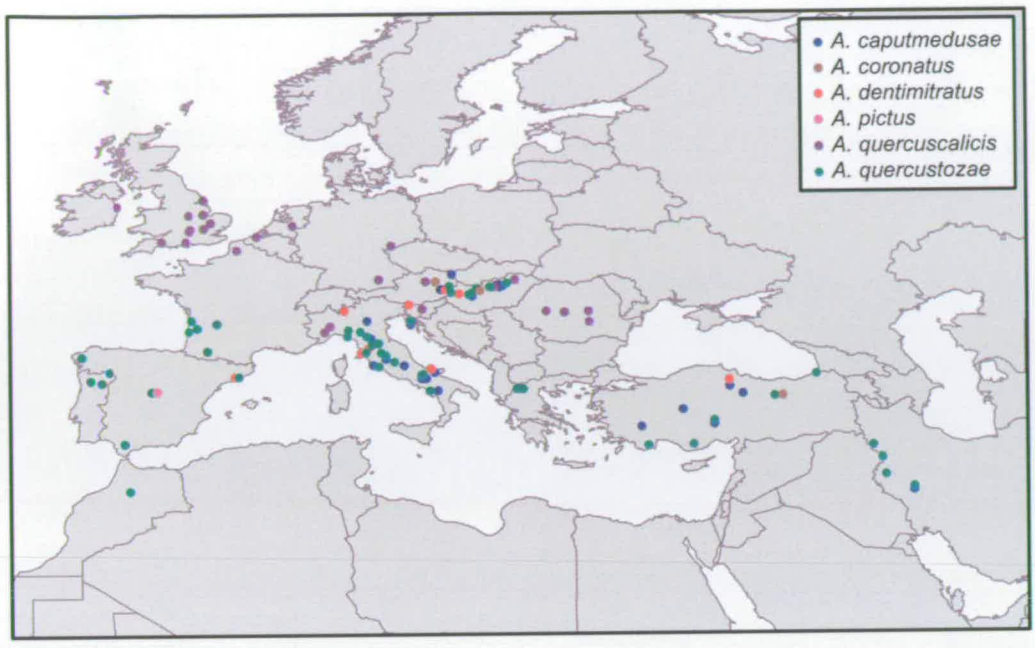


Figure 6.2 Sample locations for each of the six species represented by more than one sample. Full sample location details are presented in Appendix 6.

6.2.2 Molecular methods

DNA was extracted using the DNeasy Tissue kit (Section 2.11.2). 433 bp fragments of the mitochondrial cytochrome *b* (*cytb*) gene (Sections 2.11.3.1) and 489 bp fragments of the nuclear ribosomal 28S gene D2 region (28SD2; Section 2.11.3.3) were amplified and sequenced in both directions (Section 2.11.4). Previously published *cytb* sequences are also included for *A. askewi* (1), *A. caputmedusae* (3), *A. coronatus* (3), *A. dentimitratus* (3), *A. glutinosus* (1), *A. mitratus* (1), *A. pictus* (3), *A. quercuscalicis* (1) and *A. quercustozae* (30 from 48 individuals) (see Appendix 6 for accession numbers).

6.2.3 Sequence analysis

Since the species within the *A. quercuscalicis* clade have been inferred to be non-monophyletic on the basis of mtDNA (Rokas *et al.* 2003b), haplotypes for all species within the clade were treated as a single set and assigned to molecular operational taxonomic units (MOTUs; Section 2.10). MOTUs were defined using

MOTU_define 2.04 (Section 2.12.9). In order to define sub-clades for this study, the procedure was run with the cut-off between MOTUs set at each interval between 4 to 13 bp intervals, equivalent to between 1 and 3% sequence divergence. *G*-tests (Sokal & Rohlf 1981) were performed to test for association between MOTUs and both species and regions. Four regions were defined: (i) Iran; (ii) Turkey; (iii) Iberia (comprising Spain, Portugal, Morocco and, for *A. quercustozae* southwest France within the native range of *Q. suber*); and (iv) Europe (comprising all other countries, including France for *A. quercuscalicis*).

Model selection was performed on the *cytb* data using ModelTest 3.1 (Section 2.12.8) to select a single best model for the alignment under the Akaike information criterion, AIC, (Section 2.3.3.1). A phylogenetic network was generated in SplitsTree 4.4 (Section 2.12.13) using the selected model for the distances transformation.

6.3 Results

6.3.1 Haplotype diversity

The 256 samples yielded 174 *cytb* haplotypes, summarised in Appendix 6. Of these, four haplotypes (haplotypes 1, 2, 8 and 55) were shared between individuals of separate species. One haplotype (haplotype 8) was shared between individuals from separate geographic regions. Preliminary analysis revealed that three haplotypes (haplotypes 11, 27 and 28) did not come from species in the *A. quercuscalicis* clade, so these samples were excluded from subsequent analysis. During sample collection, species were identified on the basis of gall morphology, which may have been misleading. If wasps of two species oviposit in the same place on the host oak, chimeric galls can be formed with the appearance of one species but containing the larva of the other species. Of the 80 samples sequenced for 28SD2, 79 shared a single haplotype (Appendix 6). The remaining sample had *cytb* haplotype 11 and was excluded from subsequent analysis.

6.3.2 Phylogenetic network

The model selected as the best fit for the data was the HKY85 model (Hasegawa *et al.* 1985) with gamma-distributed rate heterogeneity and a proportion of invariant sites. This model, together with estimated base frequencies (A: 0.3586, C: 0.1184, G: 0.0944, T: 0.4286), *ti/tv* ratio (1.7129), gamma shape parameter (1.2400) and proportion of invariant sites (0.5849), was used to calculate distances in phylogenetic network reconstruction. The resulting network has been presented labelled according to species (Figures 6.3 and 6.4) and location (Figure 6.5).

6.3.3 MOTU definition

Between 1 and 42 MOTUs, defined at cut-offs between 4 and 13 bp (equivalent to between 1% and 3% sequence divergence), received greater than 50% support (Table 6.1). Between 4 and 10 bp, MOTUs at lower cut-offs were nested within MOTUs for higher cut-offs, this is illustrated for the 9 bp and 6 bp cut-offs in Figures 6.3 and 6.5, respectively. MOTU designation was relatively consistent across resamplings at cut-offs between 5 and 7 bp, with around half of the MOTUs inferred across the 100 resamplings being supported by at least 50% of the resamplings. At higher cut-offs, consistency was lower, particularly for the haplotypes assigned to MOTU B (Figure 6.3) at the 9 bp cut-off.

Table 6.1 Total numbers of unique MOTUs inferred, and numbers of MOTUs inferred by at least 50% of the random resampling orders, at each cut-off between 4 and 13 bp.

Cut-off (bp)	Sequence divergence (%)	Number of MOTUs	
		Total	≥ 50% support
4	0.92	352	26
5	1.15	59	42
6	1.39	47	23
7	1.62	37	19
8	1.85	54	13
9	2.08	58	9
10	2.31	35	7
11	2.54	219	1
12	2.78	151	1
13	3.01	152	1

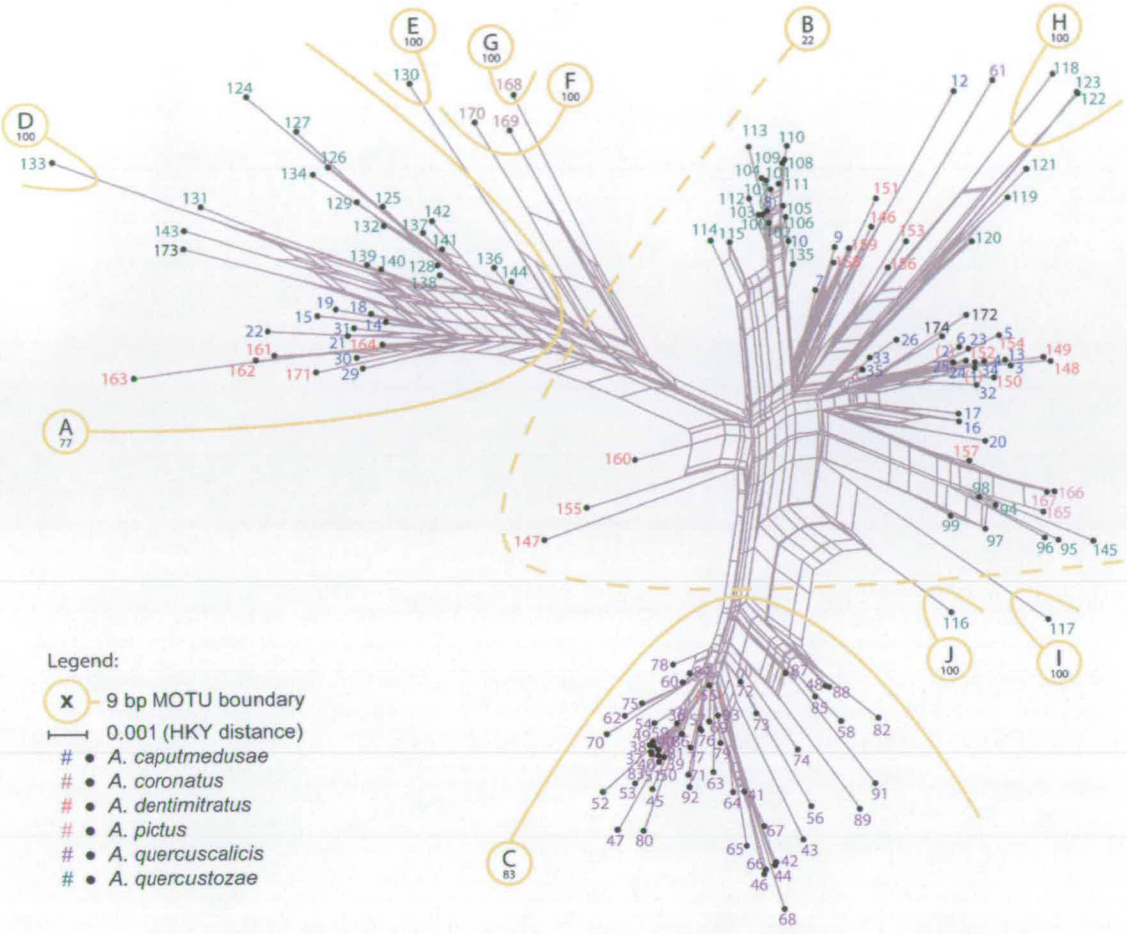


Figure 6.3 Phylogenetic network of the *A. quercuscalicis* clade produced using SplitsTree (Section 2.12.13) under the HKY distances transformation. Haplotypes are numbered as in Table 1 and shaded according to species. Parentheses are used to indicate haplotypes shared by more than one species. Haplotypes 172 (*A. mitratus*), 173 (*A. askewi*) and 174 (*A. mitratus*) are unshaded as only one sample was included for each of these species. Yellow lines indicate MOTU boundaries at the 9 bp cut-off level using MOTU_define (Section 2.12.9). Numbers in MOTU labels are the percentage of 100 random resamplings for which the labelled MOTU was recovered, dashed lines indicate MOTU boundaries with less than 50% support.

6.3.4 Concordance between MOTUs and species

Only one of the widely sampled species, *A. quercuscalicis*, is monophyletic (excluding one divergent haplotype, haplotype 61, and one haplotype shared with an *A. dentimitratus* specimen, haplotype 55). The monophyly of *A. quercuscalicis* was used as a guide to determine an appropriate cut-off between MOTUs to reflect approximately species-level divergence, although the exact level of divergence

between species will depend upon time since the most recent common ancestor (MRCA) and effective population sizes. At cut-offs ≤ 8 bp, *A. quercuscalicis* was divided into at least two MOTUs. The 9 bp level, equivalent to 2.08% sequence divergence, was the lowest level at which all *A. quercuscalicis* haplotypes were assigned to a single MOTU (MOTU C) for the majority of random resampling orders. The divisions between MOTUs at this level are shown on Figure 6.3.

The three largest 9 bp MOTU groupings, MOTUs A, B and C, were mutually exclusive across the 100 resamplings. The results of *G* tests of association between these MOTUs and species are presented in Table 6.2. Association between MOTUs and species was only significant when MOTU C (consisting entirely of *A. quercuscalicis*) was included.

Table 6.2 Results of *G tests* of association between 9 bp MOTUs and species/regions sampled.

MOTUs	Regions/species	d.f.	*G*	*P*
A, B, C	Europe*, Iberia, Iran, Turkey	6	123.16	< 0.001
A, B	Europe*, Iberia, Iran, Turkey	3	79.53	< 0.001
A, B, C	*A. caputmedusae*, *A. dentimitratus*, *A. quercuscalicis*, *A. quercustozae*	6	167.07	< 0.001
A, B	*A. caputmedusae*, *A. dentimitratus*, *A. quercustozae*	2	1.67	> 0.05

* excluding Iberia

Under an introgression hypothesis, the frequency of introgression can be estimated from the phylogenetic distribution of haplotypes for each species. Rare introgression events are expected to result in a small number of monophyletic clusters of each taxon, whereas frequent introgression events are expected to result in polyphyletic clusters. Both patterns can be seen in the *Andricus quercuscalicis* clade at the 9 bp cut-off. *A. quercustozae* haplotypes are distributed in a small number of monophyletic sub-clades distinct from the remaining species. The distribution of these sub-clades is concordant with phylogeography of the species based on allozyme data (Rokas *et al.* 2003a), demonstrating that for this species nuclear and

mitochondrial markers are concordant and any introgression has been rare. *A. caputmedusae* and *A. dentimitratus* form mixed clusters, indicating more frequent introgression between these species.

At lower cut-offs there are further single-species clusters with haplotypes from each of the four most widely sampled species in at least one single-species MOTU. At the 6 bp cut-off (Figure 6.4) *A. quercustozae* is divided into 12 clusters, however, two of these are multispecies clusters shared with *A. caputmedusae* and *A. dentimitratus* (MOTU l), or *A. dentimitratus* and *A. pictus* (MOTU s). Two further clusters contain both *A. caputmedusae* and *A. dentimitratus* (MOTUs a and q).

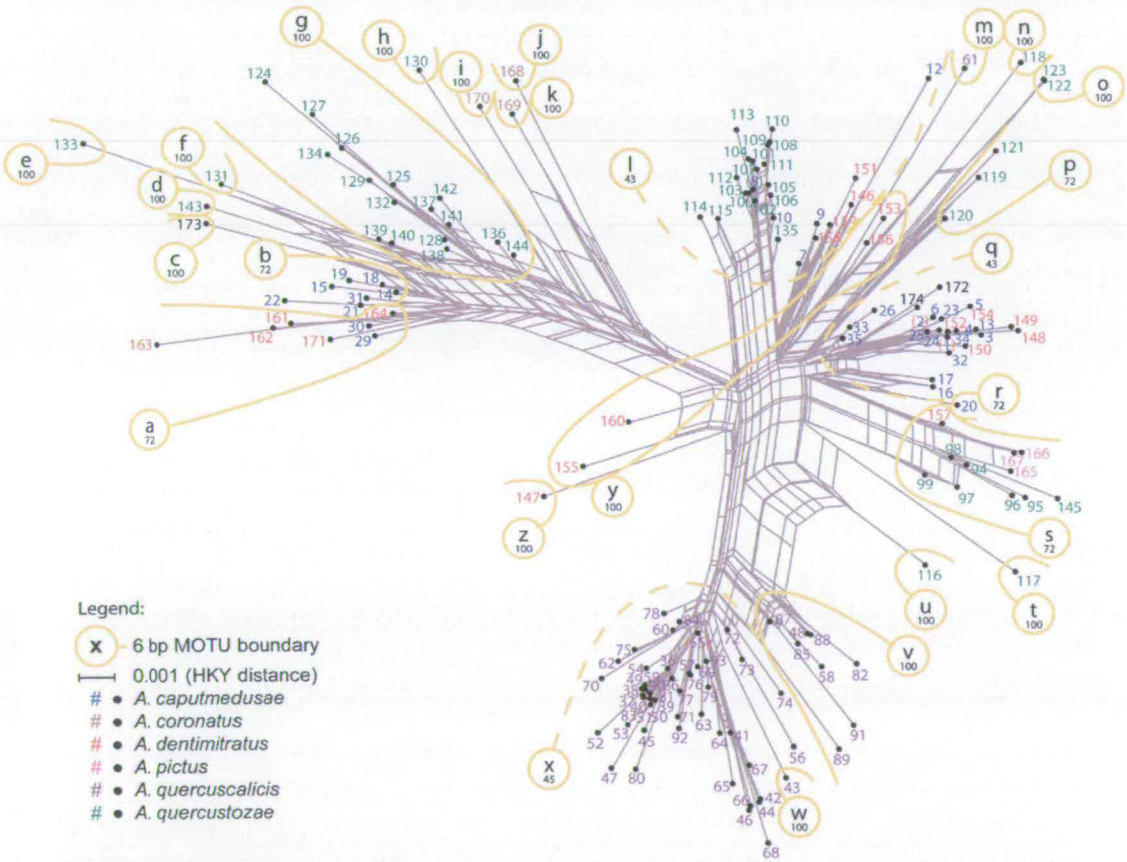


Figure 6.4 Phylogenetic network of the *A. quercuscalicis* clade produced using SplitsTree (Section 2.12.13) under the HKY distances transformation. Haplotypes are numbered as in Table 1 and shaded according to species, as in Figure 6.3. Yellow lines indicate MOTU boundaries at the 6 bp cut-off level using MOTU_define (Section 2.12.9). Numbers in MOTU labels are the percentage of 100 random resamplings for which the labelled MOTU was recovered, dashed lines indicate MOTU boundaries with less than 50% support.

6.3.5 Concordance between MOTUs and geography

There were significant associations between 9 bp MOTUs A, B and C and geographic region (Figure 6.5 and Table 6.2), regardless of whether MOTU C (*A. quercuscalicis*) was included. MOTU A contained significantly more Iranian and Turkish haplotypes while MOTU B contained more European (including Iberia) haplotypes than expected due to chance. Subdivision into 6 bp MOTUs separated some groups of Iberian Haplotypes from the rest of Europe, either patially (MOTU s) or completely (MOTU p).

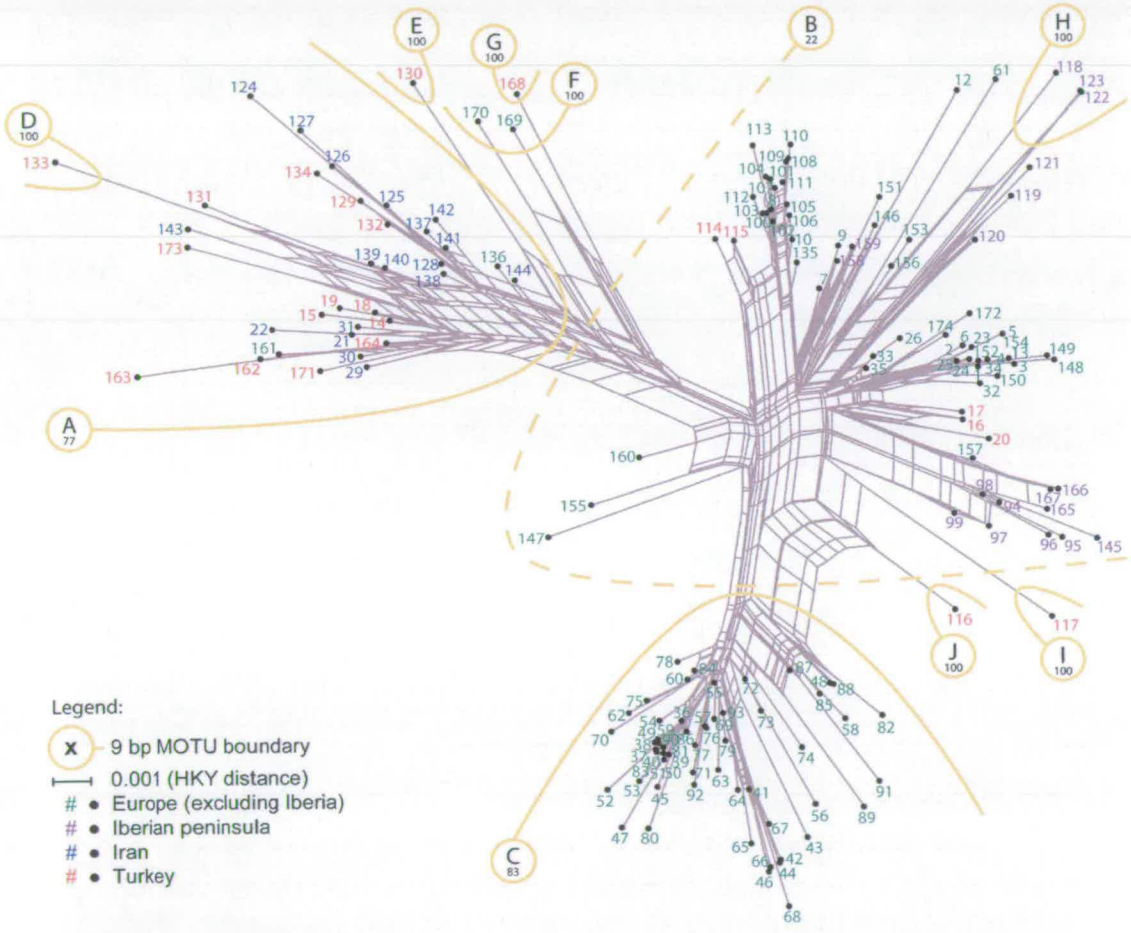


Figure 6.5 Phylogenetic network of the *A. quercuscalicis* clade produced using SplitsTree (Section 2.12.13) under the HKY distances transformation. Haplotypes are numbered as in Table 1 and shaded according to region. Parentheses are used to indicate haplotypes that were sampled in more than one region. Yellow lines indicate MOTU boundaries at the 9 bp cut-off level using MOTU_define (Section 2.12.9). Numbers in MOTU labels are the percentage of 100 random resamplings for which the labelled MOTU was recovered, dashed lines indicate MOTU boundaries with less than 50% support.

6.4 Discussion

6.4.1 Molecular concordance with morphological taxonomy and elevation of regional forms to species status

All individuals shared a single 28SD2 haplotype, so inference of phylogenetic relationships is based entirely on the mitochondrial gene *cytb*. The lack of variation in the 28SD2 gene indicates that the members of the *A. quercuscalicis* clade have a recent shared ancestry compared with both *A. coriarius* (Challis *et al.* 2007; Chapter 4) and *A. kollari* (Stone *et al.* 2007a; Chapter 5) which both showed intraspecific variation in the same gene. The phylogenetic network of the *Andricus quercuscalicis* clade is divided into three major sub-clades, which are supported by MOTU designation at the 9 bp cut-off. One of these clades (MOTU C) almost exclusively contains members of a single species, *A. quercuscalicis*, from a single region, non-Iberian Europe.

Each of the remaining large sub-clades (MOTUs A and B) contains haplotypes from more than one region and species. There is no significant association of MOTUs and species for these groups, however, haplotypes sampled from areas east of Europe (Turkey and Iran) are significantly associated with MOTU A while haplotypes from Europe (including Iberia) are significantly associated with MOTU B. This division between eastern and western lineages reflects patterns noted in individual oak gallwasp species (Challis *et al.* 2007; Stone *et al.* 2007a; Chapters 4 and 5) and wider taxonomic groups (Chapter 3).

In *A. quercustozae*, this division closely corresponds with the division between the European (MOTU B) and the *insana* (MOTU A/g) gall morphologies; only one European haplotype is found in the *insana* sub-clade. The *insana* morphology has been recorded from Greece and southern Italy, however, the European haplotype in the *insana* sub-clade was sampled in Varpolota, Hungary. This is likely to have resulted from rare hybridisation between the two gall morphologies in the Balkans, followed by backcrossing to the European morphotype. A similar pattern has been recorded in *A. kollari* where regionally atypical mtDNA

haplotypes are present in individuals with regionally typical nuclear genotypes (Stone *et al.* 2007; Chapter 5). Allozyme data support the presence of regionally distinct nuclear genotypes in *A. quercustozae* (Rokas *et al.* 2003a). The presence of a genetic division that is broadly concordant with both geography and gall morphology, albeit only for one mitochondrial gene, supports the sibling species hypothesis for *A. quercustozae* and *A. insana* (Pujade-Villar *et al.* 2002).

For both *A. caputmedusae* and *A. dentimitratus*, MOTU divisions at the 9 bp cut-off coincide with a division between European and Turkish/Iranian populations. Iranian gallwasps that were initially identified as resembling *A. dentimitratus* (Figure 6.1D) have recently been described as a new species, *Andricus assarehi* (Figure 6.1A; Melika *et al.* in prep.). No Iranian *A. assarehi* or *A. dentimitratus* were included in this study so it is not possible to determine whether the Turkish and Iranian *A. dentimitratus/assarehi* haplotypes are closely related and form a single Eastern lineage, as in *A. insana*.

Based on the limited sampling, *A. coronatus* and *A. pictus* (which is restricted to Iberia and northwest Africa) are both monophyletic. The three haplotypes of the widespread *A. coronatus* were assigned to two separate 9 bp MOTUs (E and F) and three separate 6 bp MOTUs (MOTUs i, j and k), whereas the three Iberian *A. pictus* haplotypes were assigned to MOTU B (6 bp MOTU s). Haplotypes of each of the widely sampled species were non-monophyletic and were assigned to more than one MOTU. One *A. quercuscalicis* haplotype, haplotype 61, was assigned to the MOTU B cluster. This haplotype was not closely related to any other sample and was separated into a single-species MOTU (MOTU m) at the 6 bp cut-off); however, the adult morphology could not be distinguished from *A. quercuscalicis*.

For the remaining species, polyphyly was more extensive, indicating either retained ancestral gall morphology within regions or intrarefugial introgression. Accurate distinction between these alternative hypotheses will require a more informative nuclear marker. A suitable candidate marker for this may be long-

wavelength opsin, LWRh. This marker has been used to reconstruct phylogenetic relationships among European oak gallwasp species (Hernandez-Lopez *et al.* 2007) and offers greater intraspecific resolution than the 28SD2 marker used in this study. Two alternate sources of information on the nuclear relationships among members of this clade are available: adult gallwasp phenotypes and allozyme allele frequency data for *A. quercustozae* (Rokas *et al.* 2003a). Given the discordance between mtDNA data and both morphological taxonomy and allele frequency data, and the correlation between mtDNA taxonomy and geography, intrarefugial introgression is perhaps the most likely hypothesis.

6.4.2 Within and between lineage variation in introgression

Horizontal transfer of mtDNA haplotypes can occur when reproductive barriers between species are incomplete. Introgression of the nuclear genome of one species by the mitochondrial genome of a second species has been demonstrated in several insects (e.g. Powell 1983; Thelwell *et al.* 2000; Sota *et al.* 2001). Within the oak gallwasps, rearing experiments have demonstrated that two members of the *Andricus kollari* clade (*A. kollari* and *A. lignicolus*) can mate and produce offspring capable of back-crossing to both parents for up to three generations (Folliot 1964). Hybridisation among more closely related species, such as within the *A. quercuscalicis* clade, may be possible under natural conditions (Rokas *et al.* 2003b) as has been demonstrated for *A. kollari* (Stone *et al.* 2001).

Mating in closely related *Andricus* species takes place on the gall (Stone & Sunnucks 1993; Stone *et al.* 2001) so hybridisation between lineages is likely to be most frequent where the galls of two closely related species are at high density, when both species induce galls on the same host species and when both species induce galls on the same host organ. *A. caputmedusae* and *A. dentimitratus* form a number of mixed clusters, indicating relatively frequent introgression, suggesting that both species may induce sexual generation galls on the same host organ. *A. quercuscalicis* shows very little evidence for introgression and may induce sexual generation galls on a different host organ, separating it spatially from other members of the *A.*

quercuscalicis clade, or may have different phenology. The role of temporal isolation due to differing phenology as a barrier to reproduction has been demonstrated for *A. kollari*/*A. hispanicus* (Stone *et al.* 2001).

6.4.3 Limitations of the MOTU approach in the *A. quercuscalicis* clade

Discordance between molecular and morphological taxonomy has be attributed to intrarefugial introgression, however, the inability to resolve some clusters may be due to limitations of the MOTU approach at the taxonomic level of this study:

(i) Sequence clusters can only be resolved if they exist. The members of the *A. quercuscalicis* clade are more closely related than members of comparable clades, such as the *A. kollari* clade, as all individuals shared a single 28SD2 haplotype. As such, there has been limited time for extinction of intermediate haplotypes to occur. The phylogenetic network shows a number of apparent clades with some concordance between species and geography. The presence of basal haplotypes with short terminal edge lengths reduces the genetic distinction between clades.
(ii) The definition of an appropriate cut-off based on divergence within a single species does not take variation in species' ages and effective population sizes.

(iii) Susceptibility to variation of the resampling order increases as the MOTU cut-off approaches the maximum pairwise sequence divergence in the data. As the MOTU cut-off is increased to encompass multiple clusters defined at lower cut-off levels, the influence of the choice of starting sequence on MOTU clustering of the entire dataset increases (Figure 6.6).

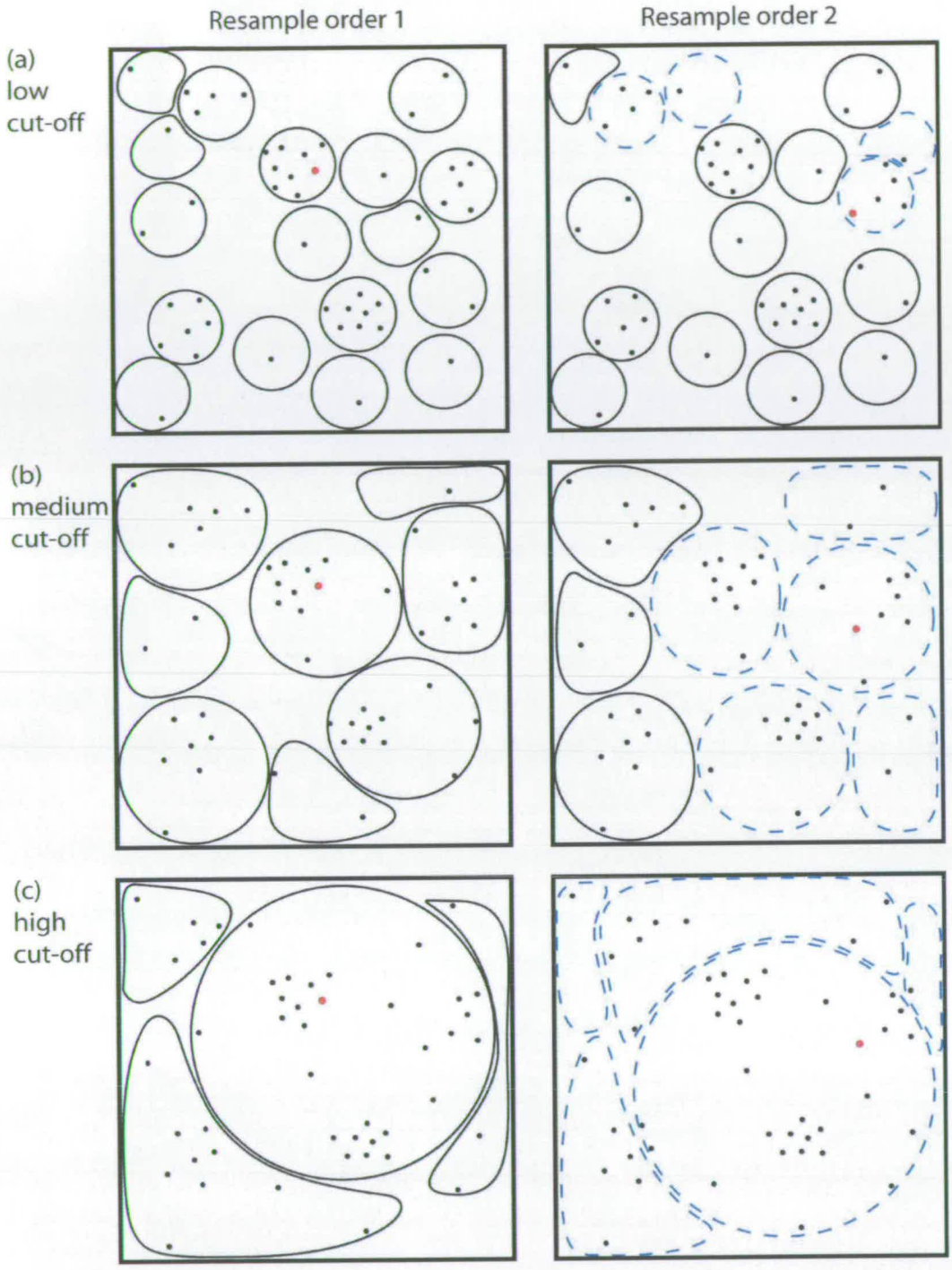


Figure 6.6 Hypothetical example of the effect of random resampling on MOTU clustering at a range of cut-offs. Each dot represents a unique haplotype and the distribution of dots reflects sequence divergence. Red dots indicate the first haplotype added to the MOTU database. Black lines delimit MOTUs defined in the first resampling order. Broken blue lines indicate novel MOTU clusters identified in the second resampling order. Resampling order has only local effects at low MOTU cut-offs. As the degree of divergence within MOTUs becomes closer to the overall level of divergence within the set of sequences, changing the starting point for resampling has a greater effect on MOTU clustering throughout the sequence set.

A combination of introgression and the issues highlighted above may affect the ability to resolve distinct MOTUs in the *A. quercuscalicis* clade.

6.5 Conclusions

The mitochondrial *cytb* data presented in this chapter suggest division of the four most widely sampled members of the *A. quercuscalicis* clade into one currently recognised species, *A. quercuscalicis*, and two regional lineages, eastern and European. The geographic divisions also correspond to divisions in gall morphology, although only to one of the previously identified putative sibling species divisions. The extent of regional speciation cannot be fully assessed without sequence data for a more informative nuclear marker, such as long-wavelength opsin. The presence of similar gall morphologies in two geographically distinct clades could reflect either intrarefugial introgression or retained ancestral gall morphology. A definitive distinction between these hypotheses will also require further nuclear sequence data. However, the retained ancestral gall morphlogy hypothesis would require assortative mating of gallwasps inducing three separate gall morphologies, which is less probable than introgression within a single closely related clade of oak gallwasps.

Chapter 7

Comparative phylogeography of Western Palaearctic oak gallwasps

Chapter 3 posed the question of whether there were shared longitudinal phylogeographic patterns across taxa in the Western Palaearctic and identified some general trends that were supported by the results of individual phylogeographic studies in Chapters 4, 5 and 6. This chapter presents a formal comparative approach to longitudinal phylogeography across a number of oak gallwasps.

7.1 Introduction

Phylogeographic studies of individual species typically investigate latitudinal post-glacial expansion routes following the last ice age (Hewitt 1999, 2004) and are important tools in determining the patterns and causes of geographic variation in the distribution of lineages (Avise 2000). For many widespread species, these Holocene range expansions repeat a pattern of range expansion and contraction that may have been occurring throughout the glacial cycles of the Pleistocene. Viewed from a longitudinal perspective, the glacial refugia that are considered as museums (Chown & Gaston 2000) of preserved diversity in latitudinal phylogeography may contain information to allow reconstruction of longitudinal colonisation from a species' origin. Reconstructing longitudinal phylogeographic processes can provide an understanding of the relationships between populations and patterns in genetic diversity across the full distribution of widespread taxa (Ludt *et al.* 2004; Challis *et al.* 2007). Comparison of longitudinal patterns across species has the potential to allow reconstruction of relationships between guilds and communities of species in linked sets of cradles (Chown & Gaston 2000) and museums, to determine whether communities have dispersed as a discrete entity or by local recruitment (Section 1.2.3).

The first aim of this chapter is to apply the tree-based protocol described in Chapter 3 to compare the phylogeographic distributions of five oak gallwasp species to investigate the potential for concordant longitudinal patterns of range expansion. Phylogeographic studies of two host-alternating (*Andricus lucidus*, including the Iranian sibling species *Andricus megalucidus*, and *Andricus grossulariae*) and two non-alternating (*Biorhiza pallida* and *Cynips quercusfolii*) species are presented together with an expanded analysis of the phylogeography of a further host-alternating species, *Andricus coriarius* (including the cryptic species *Andricus coriariformis* and *Andricus libani*; Challis *et al.* 2007; Melika *et al.* 2007; Chapter 4). While all species share the common restriction of the same asexual generation host distribution, the host-alternators may have been further restricted by the requirement for two hosts. Comparison between host-alternating and non-alternating taxa will provide an indication of the generality of shared trends across taxa with differing life histories. Each species has a broad distribution, and was sampled from its entire range (Figure 7.1; Appendix 7), extending from the Iberian Peninsula (except *A. lucidus* and *C. quercusfolii*, which are absent from this region) in the west across southern Europe to Turkey and Iran in the east. Comparisons are also drawn with the phylogeography of the host-alternating species *Andricus kollari*, whose distribution has been affected by trade (Stone *et al.* 2007; Chapter 5).

Phylogeographic comparison typically involves indirect comparison of common properties of independent analyses across two or more species (Michaux *et al.* 2005). Assessment of concordance is therefore only qualitative. The second aim of this chapter is to develop a novel concept to allow direct comparison of phylogeography across multiple species and to discuss the potential difficulties in implementing such a method. Direct comparison will allow objective assessment of phylogeographic concordance between species. Methods appropriate to direct assessment of phylogeographic concordance should also have applications in other comparisons of intraspecific relationships.

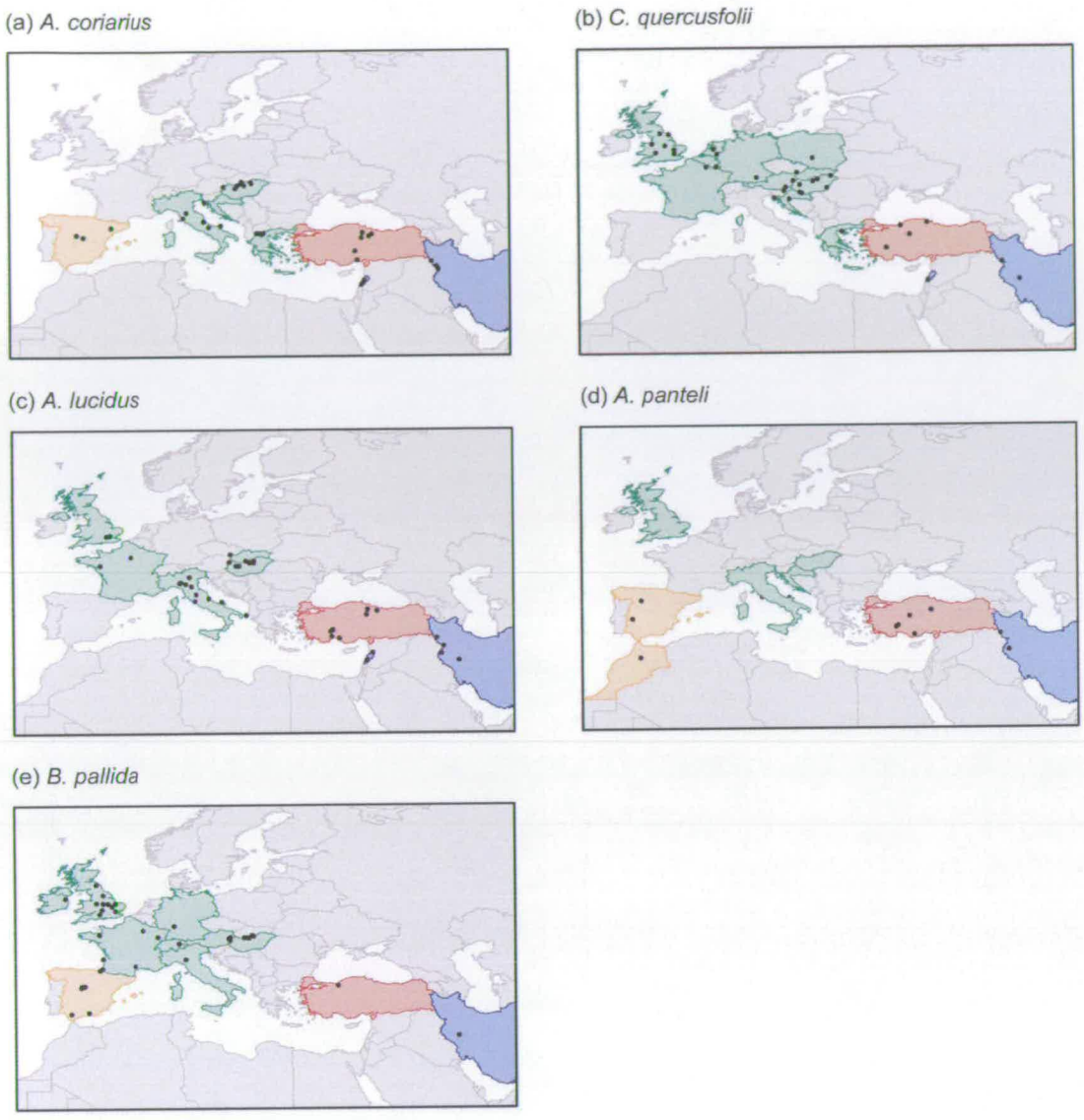


Figure 7.1 Geographical distribution of the samples of: (a) *Andricus coriarius*, (b) *Cynips quercusfolii*, (c) *A. lucidus*, (d) *A. grossulariae*, and (e) *Biorhiza pallida*. Countries from which samples were collected are highlighted according to the regional coding used in this study: ■ the Iberian Peninsula and Morocco, ■ non-Iberian Europe west of Turkey, ■ Turkey, ■ Iran and Lebanon.

7.2 Methods

7.2.1 Sample collection

Sexual and parthenogenetic generation galls of *A. lucidus*, *A. megalucidus*, *A. grossulariae*, *B. pallida* and *C. quercusfolii* were collected from locations across the species' ranges from Iberia to Iran (Figure 7.1; Appendix 7). As far as possible,

sample sizes were balanced across distribution regions (defined here as: (i) the Middle East, including Iran and Lebanon; (ii) Turkey; (iii) Europe, excluding Iberia; and (iv) Iberia). The parthenogenetic females that emerge from single multilocular galls are commonly the offspring of a single sexual female so only a single individual was sequenced from each gall for *A. grossulariae*, *A. lucidus* and *B. pallida*.

7.2.2 DNA extraction and sequencing

DNA was extracted using the DNeasy Tissue kit (Section 2.11.2). Sequence data for a 433 bp fragment of the mitochondrial cytochrome *b* (*cytb*) gene, and for the D2 region of the nuclear 28S ribosomal array were used to infer relationships between sampled individuals and selected outgroups (Sections 2.11.3 & 2.11.4). Previously published *cytb* sequences for each species and a set of outgroup taxa were also included (accession numbers: *Andricus caputmeduse*, AJ228456; *Andricus conglomeratus*, AJ228468; *Andrius conificus*, AJ228460; *Andricus curvator*, AJ228453; *Andricus kollari*, AJ228466; *Andricus seckendorfii*, AJ228449; *Andricus solitarius*, AJ228475; *Cynips cornifex*, AF539795 AJ228479; *Cynips disticha*, AF539580; *Cynips divisa*, AJ228477; *Cynips longiventris*, AF539581 AF539582; *Cynips quercus*, AF539583 AF539584 AJ228478).

7.2.3 Model selection

For each species, model selection was performed on the *cytb* data using ModelTest 3.6 (Section 2.12.8). A single best model for each data partition (Section 2.3.2) was selected using the Akaike information criterion, AIC, (Section 2.3.3.1) for: (i) the full alignment, (ii) first and second codon positions combined (iii) first codon positions only, (iv) second codon positions only, and (v) third codon positions only.

7.2.4 Trees vs. networks revisited

Previous chapters have adopted a cautious approach to the use of phylogenetic trees in phylogeography. Networks have greater power to represent the uncertainty in

phylogenetic relationships (Section 2.7). However, for mitochondrial DNA sequence data, the departure from tree-likeness is more likely to arise through retained ancestral polymorphism or hybridisation than reticulate events such as recombination. Therefore, while networks provide an implicit representation of uncertainty in regions of the phylogeny that cannot be resolved as a bifurcating tree, the concept of tree-like evolution is not violated and it is possible to integrate over uncertainty in the true topology using Bayesian inference. Since more sophisticated methods are available to analyse phylogenetic trees (see Chapter 3), a tree framework has been adopted in this chapter.

7.2.5 Phylogenetic reconstruction

Bayesian phylogenetic reconstruction was performed using MrBayes 3.1 (Section 2.12.10). For each species, phylogenies were produced under the selected models in an unpartitioned alignment, a two-partition (first and second codon positions combined *versus* third codon positions) alignment and a three-partition (each codon position) alignment. Each analysis consisted of two independent runs of four Metropolis-coupled Markov chains (MC^3) run for 2 million iterations. Convergence was assessed through the standard deviation of split frequencies between the two runs (Section 2.5.1) and visual examination of plots of parameter values (Section 2.5.2). Parameter values were sampled from the cold chain every 1000 iterations after a burn in of 400,000 iterations. The most appropriate partitioning strategy was determined by comparison of the marginal mean log-likelihoods under each set of partitions using Bayes Factors (Section 2.3.3.4). In each case a smaller number of partitions (i.e. fewer overall model parameters) was preferred unless the Bayes factor difference in favour of the more parameter rich hypothesis was greater than 20. Nucleotide diversity, π, was calculated within well-resolved clades of each species using DnaSP 4.10 (Section 2.12.5).

The 28SD2 haplotypes were aligned using the default gap formation and extension settings in ClustalX (Section 2.12.4). Due to the low number of

polymorphic sites, the alignment was treated as a single partition. Model selection and phylogenetic reconstruction were performed as for the *cytb* data.

7.2.6 Molecular clock

Marginal likelihoods under a non-clock and a coalescent strict-clock model were estimated using MrBayes 3.1 (Section 2.12.10) and the validity of the molecular clock hypothesis (Section 2.8) was tested using Bayes Factors (Section 2.3.3.4). Where the molecular clock hypothesis was supported, dates of most recent common ancestors (MRCAs) were estimated at key nodes in the Bayesian phylogeny using BEAST 1.4 (Section 2.12.4). Divergence was calibrated using the widely applied approximation for mitochondrial DNA of 2.3% sequence divergence per million years (Brower 1994). Although the actual rate of sequence divergence within the species in this study may have differed from this value, the approximation should give an indication of timescales involved and does not affect the relative time depth of nodes (Section 2.8.4). The alignment was partitioned by codon position and the Metropolis-coupled Markov Chain Monte Carlo (MC^3) chain was run for 10 million generations under the HKY model with gamma-distributed rate heterogeneity and sampled every 1,000 generations after a burn-in of 1 million generations. All parameters and likelihood values were assessed using Tracer 1.4 (distributed with BEAST 1.4; Section 2.12.4) to estimate 95% confidence intervals and ensure a sufficient effective sample size for each parameter.

7.2.7 Geographic reconstruction

Ancestral locations of clades in the phylogenies were reconstructed using BayesMultiState (Section 2.12.2) to perform likelihood mapping on a subset of 100 trees sampled at equal intervals from the Markov chain output. Geographic reconstruction was simplified by grouping putative refugia unless there was strong support for the monophyly of a refugium or set of refugia in one or more species. Each haplotype was scored according to the region(s) in which it was sampled as: (i)

Iran/Lebanon; (ii) Turkey; (iii) Europe west of Turkey to the Gironde, France; and (iv) the Iberian Peninsula and Morrocco.

7.3 Results

7.3.1 Phylogenetic reconstruction

Bayes factor (Kass & Raftery 1995) model selection demonstrated that each species' phylogeny was compatible with a strict-clock model in which the *cytb* alignment was partitioned by codon position. Intraspecific phylogenies for each of the six species in this study are presented in Figure 7.2. Three species (*A. coriarius*, *A. lucidus* and *C. quercusfolii*) have a single, poorly-resolved clade of predominantly European haplotypes and a separate clade (or clades) of eastern (Iranian, Lebanese and Turkish) haplotypes. *B. pallida* has a well-resolved clade of Iberian haplotypes distinct from a poorly-resolved predominantly European clade. A single (of two sampled) Iranian haplotype forms an outgroup to the rest of the *B. pallida* sequences. The *A. grossulariae* phylogeography lacks distinct geographic correlation but has two deeply divergent widespread clades. *A. kollari* has two geographically distinct clades (Iberian and Northern European) while the remaining sequences form a basal polytomy (Stone *et al.* 2007).

Across the six species, most haplotypes (77%) were sampled from a single individual; this value ranged from 65% for *B. pallida* to 94% for *A. grossulariae*. Among the remaining haplotypes, all of the species have at least one haplotype that was found across a large section of the longitudinal range of the species. *A grossulariae* haplotypes 4 and 35 are distributed from Spain to Turkey and from Morocco to Iran, respectively, and *C. quercusfolii* haplotype 4 is distributed from the UK to Turkey. *C. quercusfolii* haplotype 4 and *B. pallida* haplotype 1 (distributed from Spain to Hungary) were not only widespread but also occurred at high frequency among the sampled individuals (25% and 50%, respectively).

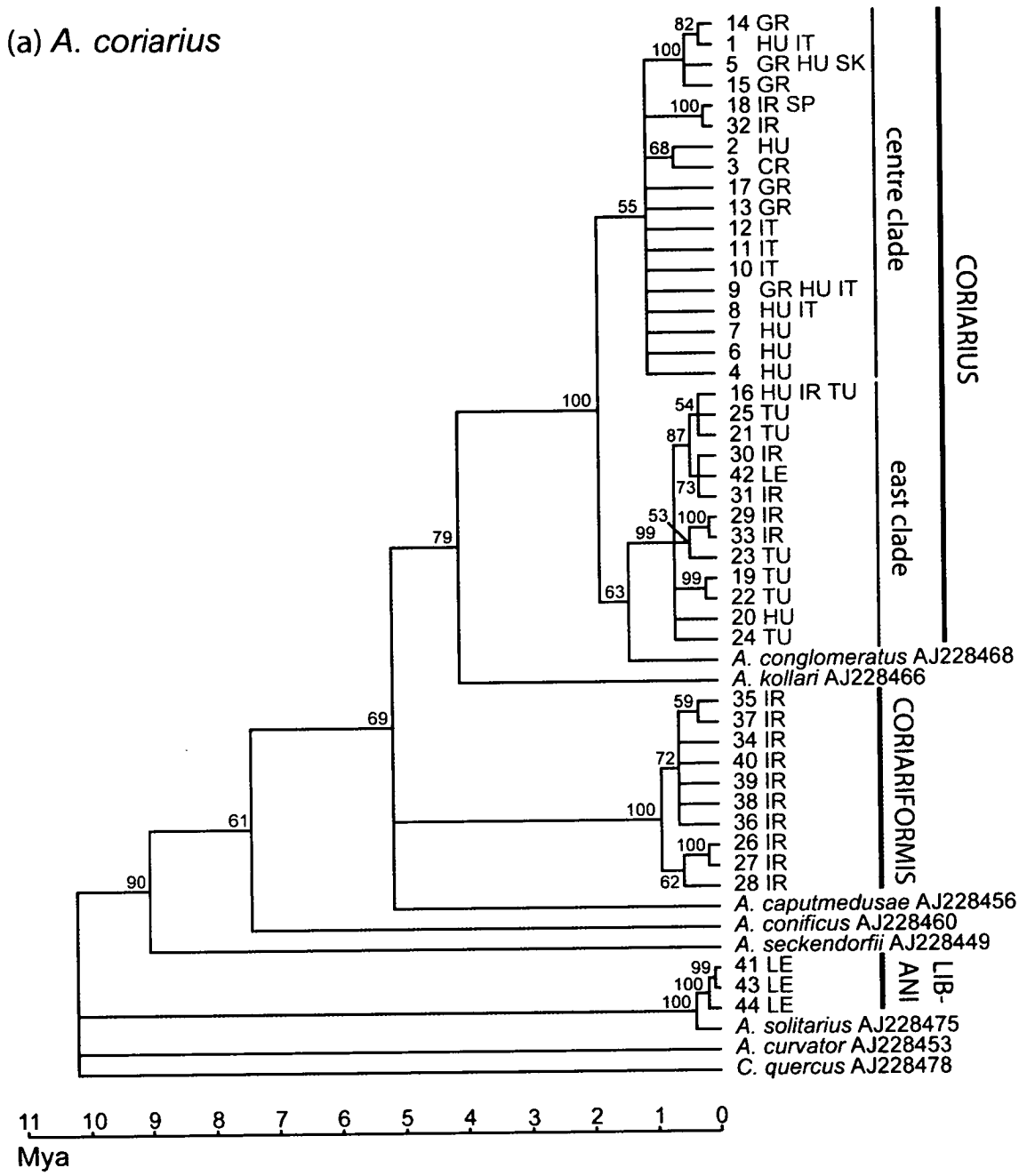


Figure 7.2 Continues overleaf.

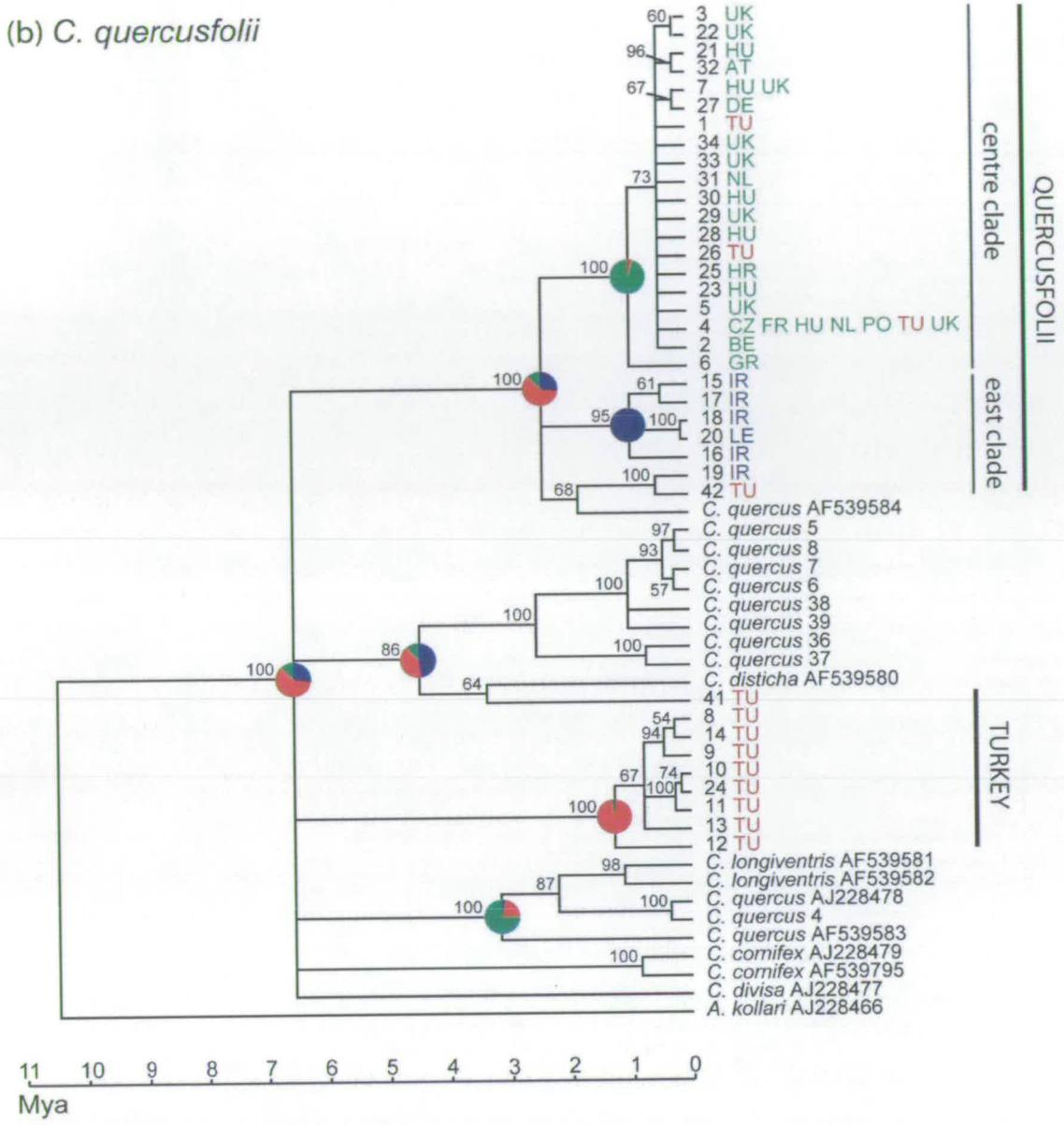
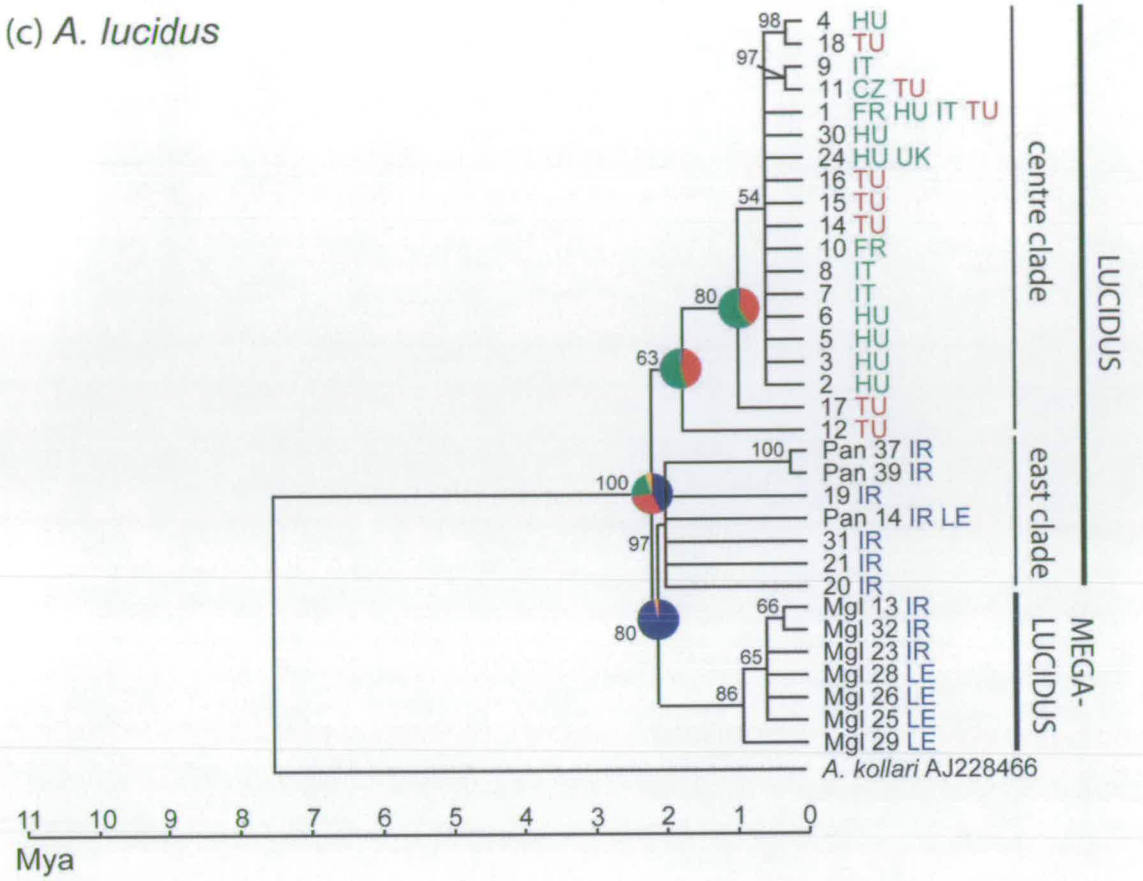


Figure 7.2 Continues overleaf.



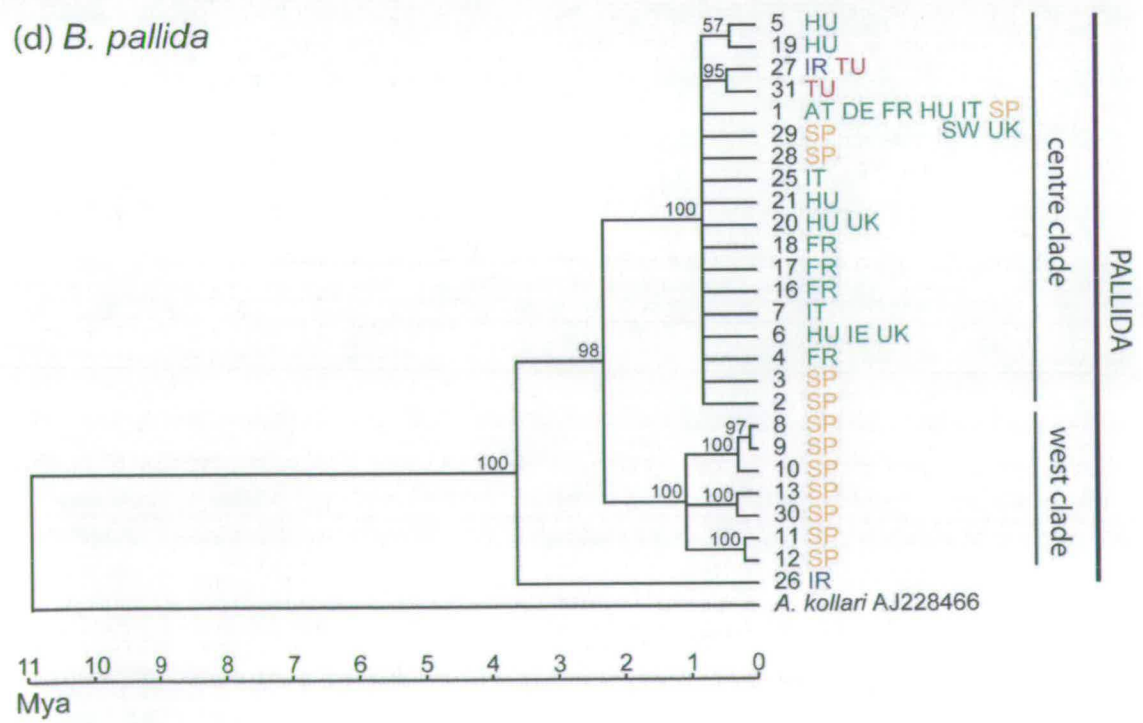


Figure 7.2 Continues overleaf.

(e) *A. grossulariae*

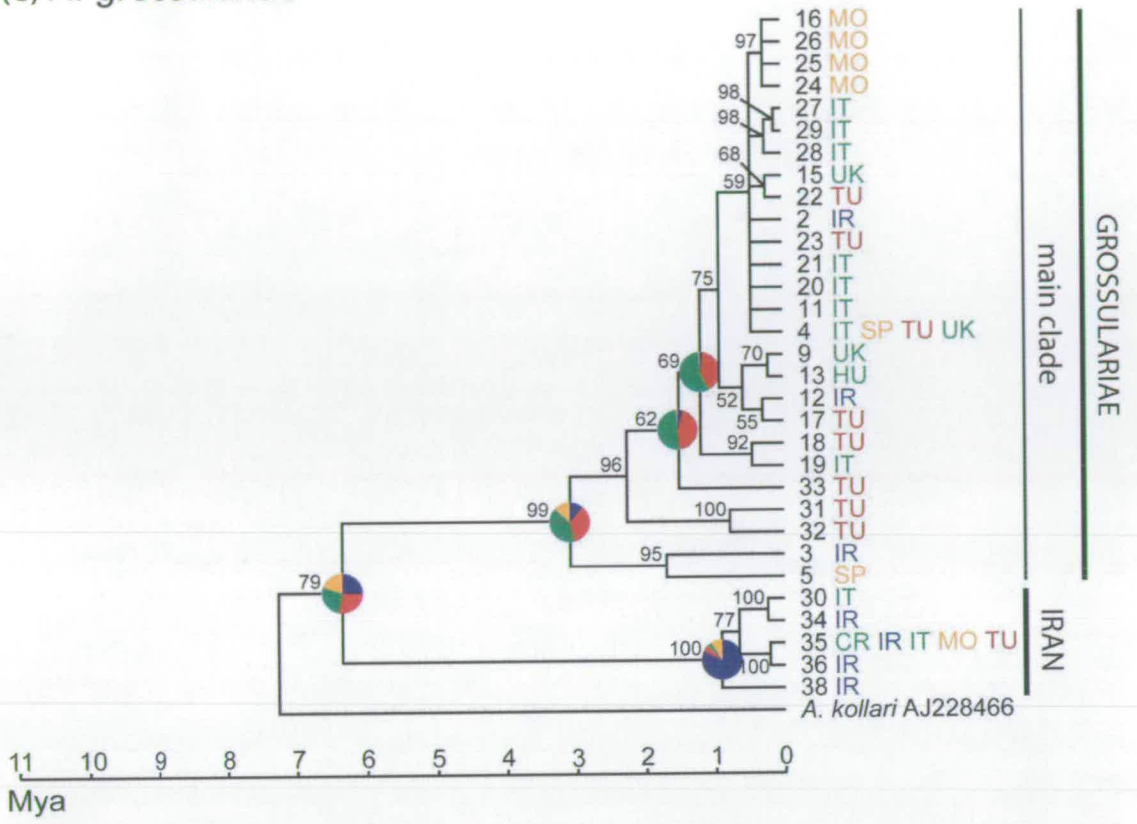


Figure 7.2 Bayesian intraspecific cytochrome *b* phylogenies of: (a) *Andricus coriarius*, (b) *Cynips quercusfolii*, (c) *A. lucidus*, (d), *Biorhiza pallida* and (e) *A. grossulariae*. All phylogenies were produced using an alignment partitioned by codon position under the coalescence strict-clock model in MrBayes 3.1 (Ronquist & Huelsenbeck 2003). Node labels are posterior probability support values. Model selection for each partition was performed using MrModeltest (Nylander 2004). Regional affiliation and countries in which haplotypes were sampled are indicated by colouring (■ the Iberian Peninsula and Morocco, ■ non-Iberian Europe west of Turkey, ■ Turkey, ■ Iran and Lebanon) and two letter codes (AT, Austria, BE, Belgium, CR, Croatia; CZ; Czech Republic DE, Germany, FR, France; GR, Greece; HU, Hungary; IE, Ireland; IR, Iran; IT, Italy; LE, Lebanon; MO, Morocco; NL, Netherlands; PO, Poland; SK, Slovakia; SP, Spain; SW, Switzerland; TU, Turkey; UK, United Kingdom), respectively. Ancestral location reconstruction was performed for *C. quercusfolii*, *A. lucidus* and *A. grossulariae* using BayesMultiState (Pagel *et al.* 2004). Pie charts indicate the relative probabilities of each of the four possible regional locations of the MRCA of a clade. Bold lines indicate nominal species-level clades.

7.3.2 Timescale

The timescale in Figure 7.2 shows the relative time depth of nodes in the five phylogenies. More detailed date estimates for the most recent common ancestors (MRCAs) and most ancient common ancestors (MACAs; Hayward & Stone 2006) (Section 2.8.3) of the main 'species-level' clades within these phylogenies, together with the MRCA estimate for *A. kollari* from Stone *et al.* (2007; Chapter 5), are presented in Table 7.1. The basal Iranian and Lebanese lineages within the *A. coriarius* phylogeny have been identified as cryptic species, *A. coriariformis* and *A. libani* (Challis *et al.* 2007; Melika *et al.* 2007; Chapter 4). Similarly divergent clades (i.e. those inferred to have diverged greater than 4 mya) within *C. quercusfolii*, haplotypes 8-14 and 24 from Hadim, Turkey, and *A. grossulariae*, haplotypes 30, 34-36 and 38, have been excluded from the estimates of species' MRCA. The MRCA for each of these species with the divergent clades included is equivalent to the MACA date of the main, 'nominal' species clade (Table 7.1). MRCA dates for the main clades of each species vary from around 1.5 mya to around 3.6 mya but, despite the broad range of MRCA estimates, only the two species at the extremes of this range, *A. kollari* and *B. pallida*, have 95% confidence intervals that are incompatible with MRCAs for all of the main species clades having occurred at the same time.

Table 7.1 Estimated most recent common ancestor (MRCA) and most ancient common ancestor (MACA) dates (Section 2.8.3) with 95% confidence intervals for the main clade of each of the species in this study. Dates were estimated from the cytochrome *b* alignments using BEAST (Section 2.12.4). Each alignment was partitioned by codon and MCMC simulations were run for 10,000,000 generations under the HKY model of evolution. MRCA estimates were sampled every 1000 generations following a burn-in of 1 million generations.

Species	MRCA date (mya)	MACA date (mya)
A. coriarius	1.8 (1.2 – 2.5)	4.0 (2.9 – 5.5)
A. kollari†	1.5 (0.9 – 2.1)	
A. lucidus	2.1 (1.2 – 3.0)	7.4 (5.1 – 9.6)
A. grossulariae	3.0 (1.9 – 4.1)	7.3 (5.5 – 9.4)
B. pallida	3.6 (2.4 – 5.6)	11.0 (7.9 – 14.5)
C. quercusfolii	2.4 (1.6 – 3.4)	6.6 (4.9 – 8.3)

† MRCA estimate from Stone *et al.* (2007); Chapter 5.

The distribution of 28SD2 haplotypes across the interspecific phylogeny (Figure 7.3) is concordant with the intraspecific distribution of *cytb* haplotypes. Each species has at least one haplotype distributed across multiple regions and the putative cryptic species in *A. grossulariae* and *C. quercusfolii*, together with *A. coriariformis*, *A. libani* and *A. megalucidus*, have distinct haplotypes from the main clades of their respective species.

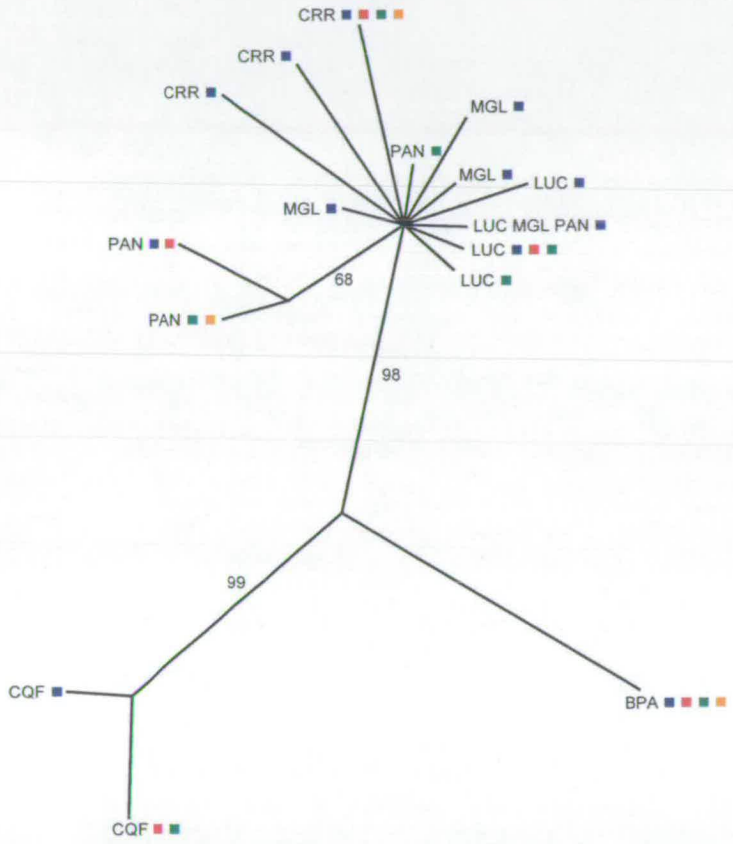


Figure 7.3 Unrooted Bayesian phylogeny of the 28SD2 haplotypes produced in MrBayes 3.1 (Ronquist & Huelsenbeck 2003) under the GTR+I+Γ model. Phylogenies were sampled every 1000 generations from two independent runs of 2 million generations, following a burn in of 400,000 generations. Branch labels are posterior probability support values. Letters are shorthand for species names: BPA, *Biorhiza pallida*; CQF, *Cynips quercusfolii*; CRR, *Andricus coriarius*; LUC, *A. lucidus*; MGL *A. megalucidus*; and PAN, *A. grossulariae*. Coloured squares indicate regional affiliation: (■ the Iberian Peninsula and Morocco, ■ non-Iberian Europe west of Turkey, ■ Turkey, ■ Iran and Lebanon.

Nucleotide divergence estimates are presented in Table 7.2 for each of the clades in Figure 7.2 that have an estimated MACA date of around 2 mya, and for the putative subspecies. Despite the common estimated date of clade origin, the π values range from 0.00209 to 0.02359 in clades in *A. lucidus* and *B. pallida* with comparable numbers of sampled individuals (60 and 76, respectively).

Table 7.2 Estimates of nucleotide divergence, π, for individual clades of each of the species in this study. Values were calculated using DnaSP (Section 2.12.5).

Species	Clade	π
A. coriarius	Centre	0.01262
	East	0.00639
	Iran	0.00795
	Lebanon	0.00924
A. kollari†	North	0.01442
	West	0.01015
	Centre/East	0.01028
A. lucidus	Centre	0.00209
	East	0.00256
	Megalucidus	0.00884
A. grossulariae	All	0.01007
	Iran	0.00665
B. pallida	Centre	0.02359
	West	0.01309
C. quercusfolii	Centre	0.00364
	East	0.00735
	Turkey	0.00903

† values from Stone *et al.* (2007); Chapter 5.

7.4 Discussion

7.4.1 Rapid dispersal

The presence of widespread mtDNA haplotypes, sampled across the entire longitudinal range of this study, suggests that the oak gallwasps are capable of rapid dispersal (since a single haplotype is maintained without mutational breakdown) on a geological timescale and of crossing many of the traditional refugial boundaries (Section 1.2.3). This inference is supported by documented rapid spread during historical gallwasp invasions in Europe (Stone & Sunnucks 1993; Stone *et al.* 2001, 2002, 2007; Csóka *et al.* 2005; Chapter 5) and associated parasitoid range expansions

(Hayward & Stone 2006), and the lack of any apparent impact of barriers such as the English Channel and Pyrenees on invasion rate. Despite this, there is phylogeographic resolution within the phylogenies for all species except *A. grossulariae*.

7.4.2 Division into eastern and western clades

Division into eastern (non-European) and western (European) clades occurred more or less simultaneously in all species studied and, according to the approximate molecular clock calibration (2.3% mtDNA divergence per million years; Brower 1994), correlates with the onset of the Pleistocene glacial cycles *ca.* 2 mya. Differentiation of monophyletic Iberian lineages was first noted in the host-alternating species *Andricus quercustozae* (Rokas *et al.* 2003) and *A. kollari* (Stone *et al.* 2001) and was attributed to the differing phenology of the Iberian (*Q. suber*) and non-Iberian (*Q. cerris*) sexual generation host, since the geographic division between clades lies to the north and east of the Pyrenees, where the distributions of these two oaks meet. A concordant pattern (albeit with a less strict dichotomy) in the non-alternating species *B. pallida* may reflect the common influence of the mountain barrier of the Pyrenees in setting up divisions between populations or trans-Pyrenean divides in section *Quercus s.s.* oak species (Stone *et al.* 2007b). However the differentiation was initially caused, it is now maintained more strongly in the host-alternators, presumably due to the more strict dichotomy of section *Cerris* oak species either side of the Gironde (Stone *et al.* 2007b) and the differing phenology of *Q. cerris* and *Q. suber* (Stone *et al.* 2001). The Iberian clades of *A. kollari* and *B. pallida* share a MACA date, once again concordant with the onset of the Pleistocene glacial cycles. This infers that oak gallwasp dispersal across the Western Palaearctic occurred in a single wave of colonisation and, except in *A. grossulariae*, the phylogeographic signature of this colonisation has not been eroded by dispersal during subsequent interglacials.

7.4.3 Common geographic origins

Inference of geographic origin was attempted for *C. quercusfolii*, *A. lucidus* and *A. grossulariae* since the distribution of sampled haplotypes for these species was considered to be relatively unbiased. While the geographic origin of *A. grossulariae* remains unresolved, the remaining two species share an inferred eastern origin. This correlates with gradients in allozyme diversity reported for *A. kollari* and *A. quercustozae* (Stone *et al.* 2001; Rokas *et al.* 2003). Challis *et al.* (2007) identified two cryptic species resembling *A. coriarius* in Iran and Lebanon. There appear to be further cryptic eastern species in the *A. grossulariae* and *C. quercusfolii* phylogenies, supported by higher 28SD2 diversity in Iran and Lebanon. These cryptic species are typically endemic, although the divergent lineage of *A. grossulariae* has colonised Europe from an Iranian ancestry, highlighting the role of the Middle East not only as a centre of diversity but also a cradle for speciation. Taken together these results demonstrate that a common geographic and temporal origin underlies the current refugial distribution of a guild of phytophagous insects.

7.5 Towards a direct comparative method

The traditional approach to phylogeographic comparison, and the approach adopted in this chapter, is to perform a set of analyses for each species independently and compare the results. This approach allows concordance in estimated dates, tree topologies and sequence summary statistics to be assessed using existing methods. However, a range of more sophisticated methodologies have been developed for comparison of phylogenies in different circumstances that have not yet been applied to phylogeography.

Comparative phylogeography, particularly among the oak gallwasps, is in many ways analogous to the coevolution of two parasites with the same host. Common climatic and geographic factors have influenced the present distributions of a number of species with shared host requirements. The main barrier to the application of host-parasite methods to comparative phylogeography is that

phylogeography is an intraspecific process. While shifts between hosts are rare in host-parasite studies, there are few genetic barriers to shift between regions and, in species with high dispersal capability, regional populations are unlikely to be genetically discrete. The challenge in developing methods for direct phylogeographic comparison is to filter out artefacts of recent dispersal to identify whether two species have concordant phylogeographic histories.

7.5.1 Artefacts of recent dispersal

Inference of a species' phylogeographic history is most difficult for clades that are composed of individuals sampled from more than one region. If the regions represented within the clade are not represented in any other clade, then there is insufficient phylogeographic resolution to separate the regions and it may be valid to combine the regions in subsequent analyses. However, if there are other clades that contain representatives of the regions concerned, then the presence of more than one region in a clade may be an artefact of recent dispersal. An intuitive approach may be to hypothesise that the region that is least represented is the recipient of recent genetic input from the region that is most represented. This hypothesis would only be valid if both regions had been sampled equally and there had been no population bottlenecks in the least represented region. An alternative hypothesis is that the least represented region has retained ancestral polymorphism and represents the source population for the most represented region. The potential for both scenarios is increased as the geographic scope of the regions to be compared is increased. Very few studies are able to collect data on the same set of species from the exact same locations so the need for relatively inclusive regions, rather than matched point locations, will be common to most applications of comparative phylogeography.

7.5.2 Assessing concordance with 'soft' labelling

Returning to the host-parasite analogy, parasite species can only occupy a single clade in a phylogeny and can therefore be coded by a single label for use in tree comparison. Due to the dispersal potential of organisms in phylogeographic studies,

individual regions can be represented in any number of clades. Labelling phylogenies with regions for use in current host-parasite methods would therefore violate the requirement for unique labelling. The use of 'soft' labelling could be introduced to allow all labels for the same region to be interpreted as equivalent (Figure 7.4). At the simplest level, soft labelled trees could then be compared using Robinson-Foulds (RF) distances (Robinson & Foulds 1981), i.e. the number of unique clades (in this case unique combinations of soft labels within clades) in a comparison of two phylogenies. The challenge in applying such a technique will be to distinguish which shared features represent a shared phylogeographic history and which are artefacts of recent dispersal. Ideally a method for direct phylogeographic comparison should also make use of information regarding the genetic diversity of clades and the frequency of occurrence of each region, in order to distinguish between recent and ancient processes.

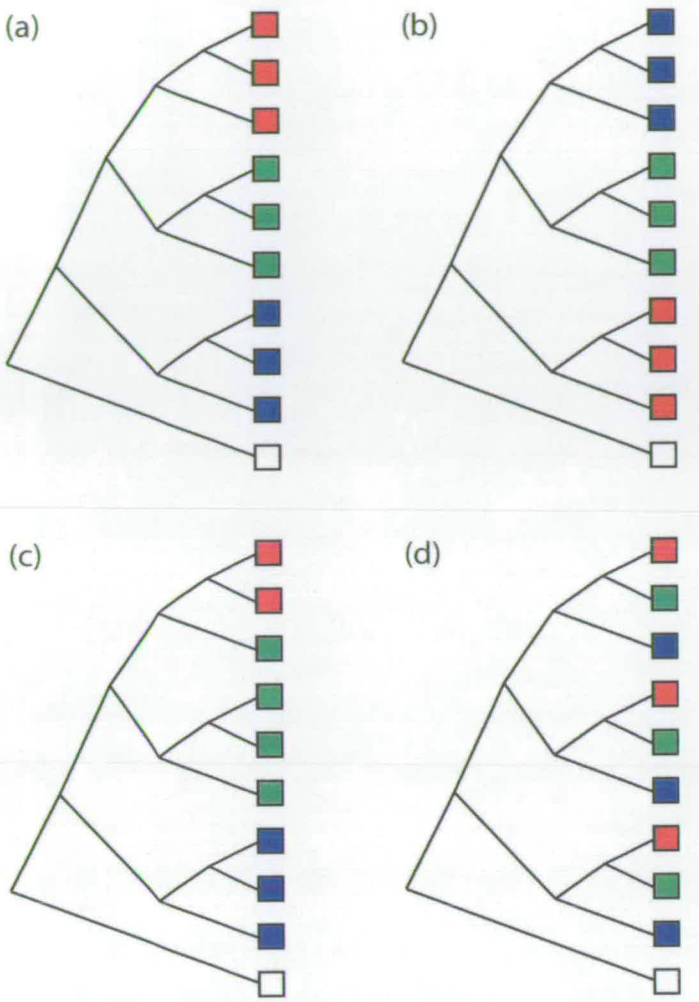


Figure 7.4 Phylogeographic comparison with 'soft' labelling. Each colour (blue, green and red) represents a 'soft' label for a separate region and white represents an outgroup. Although further features could be identified for comparison, this example focuses on two key features of phylogeny a: (i) monophyletic clades of blue, of green and of red; and (ii) blue sister group to a monophyletic clade of green and red. Each of the other phylogenies share at least one of these features: (b) monophyletic clades of blue, of green and of red, but with the red clade as a sister group to a lineage containing the other two; (c) monophyletic clades of blue and of red with blue as sister group to a monophyletic clade of green and red, and green basal to red; and (d) three separate clades in which blue is a sister group to a monophyletic clade of green and red.

Chapter 8

Phylogenetic relationships among Western Palaearctic species of the inquiline gallwasp tribe Synergini: a test of molecular barcoding and an examination of the trade-off between data quality and maximal taxon sampling

Chapters 4 to 7 have focussed on the phylogeography of the relatively well-understood oak gallwasps. This chapter addresses uncertainty in the taxonomy and phylogeny of the oak inquilines in order to provide a basis for future research into the phylogeography of this important component of oak gallwasp communities.

8.1 Introduction

Systematic biology faces the ongoing challenge of identifying and classifying millions of species. The estimated 1.5 million species that have been described (de Meeus & Renaud 2002), most of which are insects, represent only a small proportion of the estimated diversity. Molecular studies continue to reveal cryptic taxa inseparable on the basis of morphological taxonomy (e.g. Papakostas *et al.* 2005; Bergmann & Russell 2007; Starrett & Hedin 2007), increasing the magnitude of the challenge. A desire to understand ecosystem function further requires the matching of immature stages with their respective adult forms. Molecular taxonomy, and in particular DNA barcoding, promises to address these issues (Hebert *et al.* 2004b). The aim of DNA barcoding is to use sequence data for one or a few reference genes (most commonly the mitochondrial gene, cytochrome *c* oxidase subunit I, *coxI*) to assign unknown individuals to known taxa and to identify new taxa (Hebert 2004b; Powers 2004; Hajibabaei 2005; Lambert 2005; Ward 2005). Use of mitochondrial DNA (mtDNA) sequence data in phylogeographic and systematic analysis has frequently revealed cryptic diversity and/or discordance with morphology, stimulating re-evaluation of morphology-based taxonomy (e.g. Challis *et al.* 2007; Chapter 4). However, there are also potential problems associated with using mtDNA in this way. Retention of ancestral polymorphism, heteroplasmy, introgression (with

or without selective sweeps imposed by *Wolbachia* and other symbionts), potential selection on nucleotides, and the existence of nuclear pseudogenes can all lead to misleading assessment of taxonomic affinities using a single mitochondrial gene (Hebert *et al.* 2004b, Hurst & Jiggins 2005). Though apparently rare (Hebert 2003), such potential problems are best avoided by confirming assignments based on mitochondrial sequence with data for an appropriate nuclear gene (Blaxter 2004). More broadly, the utility of DNA bar-coding is arguably greatest where existing morphological taxonomic expertise allows molecular and morphological approaches to be compared, revealing the shortfalls of traditional approaches but providing biological context to sequence-based taxa (molecular operational taxonomic units, MOTUs; Section 2.10). This chapter presents a comparative approach, assessing the utility of DNA barcodes in a taxonomically challenging group of insects – the gallwasps of the inquiline tribe Synergini (Hymenoptera: Cynipidae).

The tribe Synergini (Section 1.5) contains *ca.* 170 species that inhabit plant galls induced by other insects – primarily other gallwasps in the tribes Cynipini (oak gallwasps, hosts to the inquiline genera *Ceroptres, Saphonecrus, Synergus, Synophrus* and *Ufo*), Diplolepidini (rose gallwasps, hosts to the inquiline genus *Periclistus*) and Aylacini (herb gallwasps, hosts to the inquiline genus *Synophromorpha*) (Ronquist 1994; Csóka *et al.* 2005). The major exception is the genus *Rhoophilus*, whose single species inhabits galls induced by Cecidosid moths on *Rhus* (van Noort *et al.* 2006). The inquilines are highly specialised herbivores, able to modify the host plant tissues on which they feed, but dependent on true gall inducers to initiate gall formation (Sanver & Hawkins, 2000, Stone *et al.* 2002b). Gall communities have been the subject of numerous studies of community structure and population dynamics (Stone *et al.* 2002b; Stone & Schönrogge 2003), and cynipid inquilines have a major impact on foodweb structure and community species richness (Schönrogge *et al.* 1995, 1996a, b; Schönrogge & Crawley, 2000; Stone *et al.* 2002b). However, despite extensive taxonomic work (Nieves-Aldrey & Pujade-Villar 1985, 1986; Pujade-Villar & Nieves-Aldrey 1990, 1993; Pujade-Villar *et al.*

2003), ecological studies of inquilines are severely hampered by difficulties associated with their systematics, taxonomy and identification.

The first aim of this chapter is to test the utility of *coxI* barcodes in the tribe Synergini. Since nuclear pseudogenes and introgression have both been inferred in related gall inducing Cynipini (Rokas *et al.* 2001, 2003a, b), and inquiline gallwasps are known to be infected with *Wolbachia* (Rokas *et al.* 2001, 2002a), taxon structures inferred using *coxI* are compared with those inferred from data for the D2 and D3-D5 regions of the nuclear 28S ribosomal gene. A true test of the ability of DNA barcodes to assign individuals to species requires both adequate sampling of intraspecific variation and thorough taxon sampling, including comparison of sister species (Moritz & Cicero 2004). In species-rich taxa where ambiguities exist at the genus level, all related taxa should also be sampled. This chapter focuses on the inquiline Cynipids associated with oak gallwasp hosts in the Western Palaearctic, incorporating 32 of 36 known species, to address the following questions: (a) Do *coxI* sequence data support the existing species typology, or are separate morphologically defined species represented by sequences in one MOTU (Type I error)? (b) Do DNA barcodes reveal cryptic taxa unrecognised by existing morphology-based taxonomy (Type II error)? (c) How do cytochrome *b* (*cytb*) and *coxI* compare as DNA barcode sequences? (d) Are mitochondrial and nuclear markers congruent in discrimination of taxa?

The second aim is to explore the utility of mitochondrial and nuclear gene sequence data in phylogeny reconstruction for oak-associated inquiline cynipids. The need for further phylogenetic analysis of the Synergini is shown by the fact that while morphology-based analyses uniformly support inquiline monophyly (Ronquist 1994; Ronquist & Liljeblad 2001; Nieves-Aldrey *et al.* 2005), molecular studies (Nylander *et al.* 2004a, b) suggest that inquilines are a polyphyletic assemblage comprising four distinct lineages: *Ceroptres, Synophromorpha, Periclistus,* and a *Synergus* complex (incorporating *Synergus, Synophrus* and *Saphonecrus*, with *Rhoophilus* as a sister group). This chapter focuses primarily on the *Synergus*

complex, whose genera and species are difficult to distinguish morphologically (Pujade-Villar *et al.* 2003). Though species in this complex have been included in previous molecular phylogenetic analyses (e.g. Rokas *et al.* 2002b; Nylander *et al.* 2004a, b), sampling has been very limited. This chapter uses data for fragments of the *coxI* and *cytb* mitochondrial genes and the D2 and D3-D5 regions of the 28S ribosomal gene to address the following questions: (a) Are the genera *Synophrus, Saphonecrus* and *Synergus* monophyletic groups? (b) Do sequence data support the division of *Synergus* into Sections I and II, as proposed by Mayr (Section 1.5.1)? (c) Are further significant subdivisions of *Synergus* apparent in the phylogeny?

The third aim is to explore a common dilemma facing molecular systematists: should phylogeny construction use only those taxa for which complete datasets are available (so eliminating problems associated with missing data), or should as many taxa as possible be incorporated (so reducing problems associated with missing taxa)? As DNA sequencing becomes more affordable and the availability of published sequence data increases, reconstruction of multigene phylogenies with over 100 genes or taxa is increasingly common (Bapteste *et al.* 2002; Rokas *et al.* 2003c; Philippe *et al.* 2004, 2005). The favoured approach is typically to construct a complete supermatrix in which all genes are sequenced for all included taxa (Gatesy *et al.* 2004). While this is considered the most robust approach to large-scale phylogenetic reconstruction, it has the major drawback that only taxa for which all sequence data are available can be included. This is a particular obstacle for studies based on published data as, despite calls for standardisation, the range of genes available for each varies according to the scope of the original investigation and the preferences of individual researchers. Even where all sequence data are generated specifically for the reconstruction of a large-scale phylogeny, problems such as poor storage conditions or the failure of primers in individual samples may arise which prevent construction of a complete supermatrix. In such cases, the merits of complete taxon sampling (Rannala *et al.* 1998) must be weighed against the drawback of reduced data integrity. Missing data presents a problem due to the reduction in informative sites. Given a dataset that is sufficiently large (e.g. 30,000

bp for 36 taxa used to resolve basal relationships among the Metazoa; Philippe *et al.* 2004), up to 50% missing data can be incorporated without detrimentally affecting phylogenetic inference. For smaller datasets, and particularly in studies of more recent phylogenetic relationships where factors such as retained ancestral polymorphism may reduce the correlation between phylogenetic signals at individual sites, the amount of missing data that can be accommodated is likely to be reduced. This chapter presents an empirical analysis of the integrity of an incomplete supermatrix approach as a technique to increase the breadth of taxon sampling in the phylogenetic reconstruction of the *Synergus* complex.

8.2 Methods

8.2.1 Sample collection and identification

All sampled inquilines were reared from oak galls, except *Rhoophilus loewi*, which was reared from *Scyrotis* sp. (Lepidoptera, Cecidosidae) galls induced on *Rhus* in South Africa. All known Western Palearctic species of *Ceroptres* (two species) and *Saphonecrus* (six species), two of three recognised species of *Synophrus* and 23 of 26 recognised species of *Synergus* (Csóka *et al.* 2005) were sampled. Specimens were reared from identified cynipid gall inducer hosts (Appendix 8), primarily in Hungary, a known centre of gallwasp community diversity (Stone *et al.* 2002b). To examine geographic variation within widespread morphospecies, host gall communities were also sampled in Spain and Iran. DNA was extracted from at least two specimens for all but seven species, for which only one specimen was obtained (Appendix 8). Specimens were identified by the authors of the existing keys (Pujade-Villar *et al.* 2003). However, due to difficulties in morphological identification, individual specimens have been assigned two or three morphospecies names in some cases (see Appendix 8). The Eastern Palearctic and Nearctic inquiline faunas are far less known (Csóka *et al.* 2005). This analysis incorporates three *Synergus* species from China and one from the USA.

8.2.2 DNA extraction and sequencing

184 specimens (Appendix 8), including 32 of the 36 Synergini species known from Western Palaearctic oak cynipid hosts. Typically, DNA was extracted from single hind legs using Chelex-based DNA extraction (Section Section 2.11.1), but where insect body length was less than 2 mm the whole wasp was used and DNA was extracted using a DNeasy Tissue Kit (Section 2.11.1).

Sequences were obtained for four genes (Sections 2.11.3 and 2.11.4): cytochrome *b*, *cytb*; cytochrome *c* oxidase subunit I, *coxI*; and both the D2 and D3-5 subunits of the nuclear ribosomal 28S gene, 28SD2 and 28SD35. All new sequences have been deposited in GenBank, with the accession numbers in Appendix 8. The following sequences of three inquiline species used in this study were obtained from GenBank: *Synergus crassicornis* for 28S AY368936, *coxI* AY368909; *Synergus gallaepomiformis* for 28S AF395151, *coxI* AF395175, *cytb* AF395137; *Ceroptres cerri* for 28S AY368935, *coxI* AY368910.

8.2.3 MOTU designation

Since the species-level taxonomy of the inquiline *Synergus* complex remains uncertain, the correspondence between current taxonomy and molecular sequence data was tested by defining molecular operational taxonomic units (MOTUs; Section 2.10).

In order to compare molecular and morphological taxonomic inference objectively, it is first necessary to define a criterion to determine the level of sequence difference at which to define a MOTU. Although previous barcoding studies may provide a guide to the appropriate order of magnitude of MOTU cut-offs, relevant cut-offs must be assessed *de novo* for each gene for each taxon set. Here, the sequence difference between species in the Synergini genera, *Ceroptres* and *Rhoophilus*, which have been inferred to be basal to the *Synergus* complex in previous studies (Nylander *et al.* 2004a), was used. These genera are less species

rich than the *Synergus* complex and were therefore considered to be more likely to have been accurately resolved through morphological taxonomy.

MOTUs were defined for the complete *coxI* dataset at a range of cut-off values using MOTU_define.pl (Section 2.12.9). *coxI* MOTUs were defined at cut-offs between 1 and 100 base pairs in order to determine the minimum cut-off at which the *Ceroptres* and *Rhoophilus* species were each assigned to single MOTUs and the maximum cut-off at which each species was assigned to a separate MOTU. These two cut-off values reflect the most divisive and inclusive approaches to species approximation by molecular taxonomy and are considered in order to determine the maximum and minimum number of inferred MOTUs, respectively. An intermediate approach was also adopted by defining MOTUs at a cut-off that was the midpoint between these two values.

8.2.3.1 Comparing MOTU definition across genes

An approach was developed using an analogue of Robinson-Foulds, RF, distances (Robinson & Foulds 1981) to select the three equivalent cut-offs (divisive, inclusive and intermediate) in each of the other genes. MOTU clusters were treated as clades on a phylogeny with a flat topology, consisting of the root node and the MOTU cluster nodes with the sequences as terminal nodes. RF distances were calculated as the minimum number of unique MOTU clusters between MOTU sets for *coxI* and each of the other genes. Equivalent cut-offs were defined by the following procedure: (i) *coxI* MOTUs were redefined across 100 resampling orders for a subset of sequences from individuals for which all three gene sequences were available using the three cut-offs defined above; (ii) both *cytb* and 28SD2 MOTUs were defined for the same set of individuals across 100 resampling orders at cut-offs between 1 and up to 50 base pairs, using 1 bp increments; (iii) average RF distances (across 100x100 pairwise comparisons) were then calculated for each of the three *coxI* cut-offs and each of the cut-offs for each of the other two genes. The greater the RF distance between two MOTU sets, the greater the discordance between the two genes at the cut-off under consideration. By choosing cut-offs for *cytb* and 28SD2 at

which the RF distance was minimised, comparisons could be made between the MOTUs inferred with each gene. MOTUs were also defined for the full set of sequences that were available for each of *cytb* and 28SD2 at the inclusive, exclusive and intermediate levels to determine whether samples for which sequence data were not available for all three genes could be assigned to the same MOTU as a sample for which a complete set of sequence data were available.

8.2.4 Phylogeny reconstruction

All *coxI* (660 bp) and *cytb* (433 bp) sequences were the same length, and for each fragments could be aligned unambiguously by eye. The 28S fragments were of variable length (D2 520-572 bp, D3-5 511-513 bp) and were aligned using MUSCLE 3.6 (Section 2.12.10) using default settings.

Bayesian phylogenetic inference was performed in MrBayes 3.1 (Section 2.12.9). For each data matrix, two independent Markov chain Monte Carlo (MCMC) runs of four metropolis-coupled chains were performed. Datasets were partitioned (where applicable) by gene, by codon position and/or by an approximation of 28S stem-loop regions, and a GTR+I+G model of sequence evolution was applied to each partition. While this may represent slight over-parameterisation, particularly for the three-gene dataset, the consistency of model selection simplifies comparison of the three genes and, given the efficiency of MCMC sampling in MrBayes, it was considered more appropriate to apply a model with too many parameters than too few. Phylogenies were constructed using individual genes and supermatrices of two or three genes with the gamma shape parameter, the proportion of invariant sites, base frequencies and substitution rates unlinked across all partitions. Sampling frequency, burn in time and run length were adjusted according to the behaviour of the Markov chains: a typical starting point was a run length of 2 million generations, sample frequency of 1000 generations and a burn in time of 1 million generations. Convergence was assessed using the average standard deviation of split frequencies between the two independent runs and by examination of plots of chain parameters.

All trees were rooted using *Ceroptres* species, since previous studies (Nylander *et al.* 2004a) and preliminary analyses (data not shown) have shown this genus to represent a lineage distinct from the *Synergus* complex.

8.2.5 Likelihood mapping

Likelihood mapping was performed in TreePuzzle 5.0 (Section 2.12.14) using the HKY model of nucleotide substitution. Likelihood maps were constructed for each gene with parameters estimated from the data set and using all possible quartets of taxa.

8.2.6 Tests of congruence

Overall topological congruence between the nuclear and mitochondrial genes was estimated by partitioning the dataset by gene and performing analyses: (i) with topology linked across all partitions; (ii) with topology of each partition unlinked from the other two; and (iii) with the topology of all three partitions unlinked. Marginal likelihoods of phylogenies produced under each of these conditions in MrBayes were compared using Bayes factors (BF; Section 2.3.3.4). Bayes factors were calculated as twice the difference between marginal likelihoods of two models and were interpreted according to the guidelines of Kass & Raftery (1995), with a BF over 20 taken to indicate strong evidence.

8.2.7 Data reduction

Of the 185 specimens in the study, 91 were sequenced for only one of the three genes. A further 70 specimens (including six outgroups to the Synergini) had full sequences for all three genes (*cytb*, *coxI* and 28SD2). These 70 specimens represent the **maxdata** dataset. For the remaining 24 specimens, sequence data could not be obtained for a total of 41 genes due to PCR failure. These 24 taxa were added to the maxdata supermatrix to make the **maxtaxa** dataset, which had 16.3% missing data. To assess the significance of incomplete datasets for single specimens in phylogeny reconstruction, the impact of removing data for individuals for which complete three-

gene datasets were available was examined by simulation. For simplicity, each of the three genes in the alignment of 70 taxa was assumed to contribute an equal amount of phylogenetic information, and the alignment thus represents 210 equal units of information. Random sets of genes were selected from this set and replaced with missing data, representing 5, 10, 15, 20, 25 and 30% of the total data. Three permutations were performed for each level of missing data to accommodate variation in the impact of missing data according to the specific sequences that were removed. In cases where this procedure led to the removal of all of the data for a given taxon, data for a single gene was retained at random to allow inclusion of the taxon in the analysis. Phylogenies were estimated using each of these incomplete supermatrices by the methods described above.

The phylogenetic topologies inferred using the reduced maxdata matrix were compared with the topology based on the complete maxdata matrix. For phylogenies at each level of data reduction, the likelihood under a GTR+I+G model of the complete maxdata matrix given each of the topologies sampled during Bayesian phylogenetic inference was calculated using PAUP* 4b10 (Swofford 2001). The harmonic mean of the log-likelihoods was used to represent the marginal likelihood for comparison using Bayes factors. MCMC methods simultaneously estimate the topology and the values of the model parameters. It was not possible to fix the model parameters when calculating likelihoods to compare topological differences alone since the model parameters would have varied throughout the MC^3 runs for both the complete maxdata matrix and the reduced maxdata matrices.

The resolution of the phylogenetic hypotheses generated from incomplete supermatrices was assessed to determine the impact of missing data on the inference of relationships at different depths (substitutions per site) from the tips of the phylogeny. The posterior probabilities of all nodes receiving greater than 50% posterior probability support in the complete supermatrix phylogeny were plotted against evolutionary depth from the tips of the tree, calculated as the average branch length/sum of branch lengths in the consensus treefile generated by MrBayes. This

was repeated for phylogenetic hypotheses based on incomplete supermatrices. In each case posterior probabilities were plotted against the depth of the equivalent node in the complete supermatrix phylogeny.

8.2.8 Inclusive 'maxtaxa' supermatrix

Taxa for which sequence data were not available for all three genes were handled in one of two ways: (i) if the taxon could be assigned to the same MOTU as a taxon for which all three genes had been sequenced, it was not included in the multigene phylogenetic reconstruction since its phylogenetic position could be inferred from the position of members of the same MOTU; or (ii) if the taxon was assigned to a MOTU for which no taxon had three gene sequences, one taxon from that MOTU (the taxon with the most complete sequence data) was added to the maxdata supermatrix. Phylogenetic reconstruction based on the resulting inclusive, maxtaxa, supermatrix was performed as for the maxdata supermatrix.

8.3 Results

8.3.1 MOTUs

8.3.1.1 Defining MOTUs

The divisive and inclusive cut-offs (as defined in the Methods) for the full *coxI* dataset were identified as 16 and 41 bp, respectively. An intermediate cut-off of 29 bp was also selected. Across the 660 bp *coxI* fragment, these cut-offs represent 2.4 to 6.2% sequence divergence. Equivalent MOTU cut-offs in each of the other genes (*cytb* and 28SD2) were identified using sequences from the 64 sample subset (the maxdata set excluding the 6 outgroups) for which sequence data were available for all three genes. The smallest RF distances between MOTU cluster sets identified with *cytb* and D2 and those at each selected cut-off for *coxI* are summarised in Table 8.1. Three corresponding cut-offs were identified for *cytb* (equivalent to 1.8, 4.4 and 5.1% sequence divergence, however, the 1 bp cut-off for 28SD2 (equivalent to 0.02 % sequence divergence) was most similar to the 41 bp cut-off for *coxI* so 28SD2 MOTUs could not be split further.

Table 8.1 MOTU clustering of each of the 110 *coxI* sequences at the 41 ('lumper'), 29 (intermediate) and 16 ('splitter') base-pair cut-offs. Each MOTU is separated by a solid line. The percentage of 100 random resampling orders for which the identified sequences were placed in the same MOTU is presented for each cut-off. Dashed lines indicate MOTU divisions supported by less than 100 percent of resampling orders. In such cases the value above the dashed line is the percentage of resampling orders for which all sequences between solid lines were placed in a single MOTU, the value below the dashed line is the percentage of resampling orders for which the sequences were divided into two separate MOTUs. Bold-face is used to highlight samples that were identified morphologically as belonging to the same species, but which have been assigned to separate MOTUs. Samples for which all three gene sequences were available have been shaded.

MOTU number	Morphological species identity	Cut-off (bp)		
		41	29	16
0.1	*Ceropteres cerri* #	100	100	100
	Ceropteres cerri S36			
	Ceropteres cerri S37			
0.2	*Ceroptres clavicornis* S34	100	100	100
	Ceroptres clavicornis S35			
0.3	*Ceropteres cornigera* S120	100	100	100
0.4	*Rhoophilus loewi* S163	100	100	100
	Rhoophilus loewi S164			
	Rhoophilus loewi S165			
1	***Synergus apicalis/tibialis*** S48	100	100	100
	Synergus apicalis/tibialis S52			
2	*Synergus crassicornis* #	100	100	100
	Synergus crassicornis S69			
	Synergus crassicornis S70			
	Synergus crassicornis S132			
A	*Synergus clandestinus* S57		100	100
	Synergus clandestinus S58			
3	*Synergus gallaepomiformis/pallicornis* S16	100	100	85
	Synergus gallaepomiformis/pallicornis S17			
	Synergus gallaepomiformis/pallicornis S18			
	Synergus gallaepomiformis/pallicornis S19			
	*Synergus gallaepomiformis/**pallipes*** S24			
	*Synergus gallaepomiformis/**pallipes*** S25			
	Synergus gallaepomiformis S14			
	Synergus gallaepomiformis S15			
	Synergus pallicornis S20			
	*Synergus pallicornis/**pallipes*** S22			
	*Synergus pallicornis/**pallipes*** S23			
	*Synergus pallicornis/**pallipes*** S88			
	*Synergus pallicornis/**pallipes*** S89			
	Synergus pallipes S26			
	Synergus pallipes S27			
	Synergus pallipes S86			
	Synergus pallidipennis S12			
A	*Synergus pallicornis* S21			15
	Synergus apicalis/tibialis S41			
B	***Synergus pallipes*** S87		100	100
4	***Synergus umbraculus*** S2	100	100	82
	Synergus umbraculus S100			
	Synergus umbraculus S101			
	Synergus umbraculus S102			
	Synergus umbraculus S103			
	Synergus hayneanus S81			
	Synergus sp. S109			
A	***Synergus umbraculus*** S1			18
B	***Synergus flavipes*** S38			100
	Synergus umbraculus S3			
5	*Synergus **hayneanus**/reinhardi* S6	100	100	100
	*Synergus **hayneanus**/reinhardi* S7			

MOTU number	Morphological species identity	Cut-off (bp)		
		41	29	16
5 cont.	*Synergus* ***hayneanus****/reinhardi* S75			
	Synergus ***hayneanus****/reinhardi* S76			
	Synergus hayneanus S9			
	Synergus hayneanus S73			
	Synergus hayneanus S77			
6	*Synergus consobrinus* S44	52	100	100
	Synergus consobrinus S45			
	Synergus consobrinus S55			
	Synergus consobrinus S56			
	Synergus thaumacerus S42			
	Synergus thaumacerus S43			
A	***Synergus pallipes*** S83	48	100	100
7	***Synergus hayneanus*** S78	100	100	66
	Synergus hayneanus S79			
A	***Synergus hayneanus*** S80			44
8	*Synergus variabilis* S99	100	100	100
	Synergus flavipes S39			
	Synergus flavipes S130			
	Synergus flavipes S131			
A	*Synergus acsi* S112			100
	Synergus palmirae S128			
9	***Synergus hayneanus*** S8	100	100	100
	Synergus hayneanus S74			
	Synergus hayneanus/umbraculus S10			
	Synergus hayneanus/umbraculus S11			
10	*Synergus physocerus* S28	100	100	100
	Synergus physocerus S29			
	Synergus physocerus S60			
	Synergus physocerus S61			
11	***Synergus pallidipennis*** S13	100	100	100
	Synergus diaphanus S30			
	Synergus diaphanus S31			
12	*Synergus thaumacerus* S53	100	100	100
	Synergus thaumacerus S54			
	Synergus thaumacerus S138			
13	*Saphonecrus lusitanicus* S66	72	100	100
	Saphonecrus lusitanicus S67			
A	*Saphonecrus lusitanicus* S68	28	100	100
14	*Synergus mikoi* S105	100	100	100
	Synergus mikoi S106			
15	***Synophrus politus*** S134	100	100	100
	Synophrus politus S135			
16	***Synergus pallipes*** S62	100	100	100
	Synergus pallipes S63			
17	*Saphonecrus undulatus* S46	100	100	100
	Saphonecrus undulatus S47			
18	*Synergus plagiotrochi* S64	100	100	100
	Synergus plagiotrochi S65			
19	***Synergus apicalis/tibialis*** S40	100	100	100
	Synergus apicalis/tibialis S51			
20	*Synergus japonicus* S92	100	100	100
	Synergus japonicus S96			
21	*Synergus chinensis* S90	100	100	100
22	*Synergus xiaolongmeni* S94	100	100	100
23	*Synergus incrassatus* S59	100	100	100
24	*Saphonecrus connatus* S50	100	100	100
25	*Saphonecrus haimi* S49	100	100	100
26	*Synergus gallaepomiformis* #	100	100	100
27	***Synophrus politus*** S32	100	100	100
28	*Synophrus pilulae* S33	100	100	100

8.3.1.2 Correlation between *coxI* MOTUs and morphology

The species of the *Synergus* complex (excluding *Ceroptres* and *Rhoophilus*) were assigned to 28 MOTUs at the 41 bp cut-off (Table 8.1). These MOTUs comprised 20 clusters and eight singletons. A further two singleton divisions (MOTUs 6A and 13A) receive support from less than 50% of the random resampling orders. Eight MOTUs contain samples identified morphologically as belonging to more than one species (Type I error) and seven species are represented by samples in more than one MOTU (Type II error). At the 16 bp cut-off the species of the *Synergus* complex have been assigned to 34 MOTUs, comprising 23 clusters and 11 singletons, with one further cluster and two further singletons supported by less than 50% of the random resampling orders. Nine MOTUs showed Type I error and eight species showed Type II error.

8.3.1.3 Correlation between MOTUs defined using different genes

Correlation between MOTUs defined by the two mitochondrial genes is greater at the inclusive and intermediate cut-offs (RF ~ 13; Table 8.2) than at the divisive cut-off (RF = 18.2; Table 8.2). Even accounting for the difference in length between the *coxI* (660 bp) and the *cytb* (433 bp) fragments, the equivalent cut-offs in *cytb* have lower sequence divergence than in *COI*. This is presumably the result of differing patterns of sequence conservation/selection acting on the two fragments. At the inclusive cut-off (41 bp for *coxI*, 22 bp for *cytb*), 8/14 clusters and 12/13 singletons inferred with the 64 *coxI* sequences were also inferred with *cytb* (Table 8.3). Subdivisions of the *cytb* MOTUs at the 19 bp intermediate and 8 bp divisive levels provided support for one further cluster (Table 8.3). Comparison between *coxI* and 28SD2 at the inclusive cut-off (41 bp for *coxI*, 1 bp for 28SD2) gave the lowest RF distance of all between gene comparisons (8.6; Table 8.2). The 28SD2 data supports 10/14 of the *coxI* clusters and all of the *coxI* singletons (Table 8.4). There is weak support for a further two *coxI* clusters.

Table 8.2 Robinson-Foulds (RF) distances, with standard deviations across 100 by 100 pairwise comparisons, between MOTU cluster sets identified for *coxI* at the 41, 29 and 16 bp cut-offs and the closest matching cut-off for *cytb* and 28SD2. All MOTU definitions were based on the 64 sample subset for which sequence data were available for all three genes.

***coxI* cut-off (bp)**	***cytb* cut-off (bp)**			**28SD2 1 bp cut-off**
	22	**19**	**8**	
41	**13.3 (± 1.6)**	13.9 (± 1.5)	20.8 (± 1.7)	**8.6 (± 2.5)**
29	15.1 (± 1.6)	**13.0 (± 0.0)**	18.5 (± 1.5)	10.9 (± 2.0)
16	20.4 (± 1.9)	18.7 (± 1.5)	**18.2 (± 2.1)**	17.4 (± 2.1)

Table 8.3 MOTU clustering for *coxI* (41 bp cut-off) and cytb (22, 19 and 8 bp cut-offs) sequences from the set of samples for which sequences were available for all genes in this study. Each MOTU is separated by a solid line. The percentage of 100 random resampling orders for which the identified sequences were placed in the same MOTU is presented for each cut-off. Dashed lines indicate MOTU divisions supported by less than 100 percent of resampling orders. In such cases the value above the dashed line is the percentage of resampling orders for which all sequences between solid lines were placed in a single MOTU, the value below the dashed line is the percentage of resampling orders for which the sequences were divided into two separate MOTUs. Samples for which all three gene sequences were available have been shaded.

MOTU number	**Morphological species identity**	**Cut-off (bp)**			
		COI	***cytb***		
		41bp	**22bp**	**19bp**	**8bp**
0.1	*Ceroptres cerri* S36	100	100	100	100
	Ceroptres cerri S37				
	Ceroptres cerri S166				
0.2	*Ceroptres clavicornis* S34	100	100	100	100
	Ceroptres clavicornis S35				
0.3	*Ceroptres cornigera* S120	100	100	100	100
0.4	*Rhoophilus loewi* S164	100	100	100	100
	Rhoophilus loewi S165				
1	*Synergus apicalis/tibialis* S48	100	100	100	31
	Synergus apicalis/tibialis S52				
2	*Synergus crassicornis* S69	100			69
	Synergus clandestinus S58		100	100	100
	Inquiline *sp.* S203				
	Inquiline *sp.* S204			100	100
	Synergus umbraculus S139				
	Synergus umbraculus S140				
	Synergus umbraculus S141				
	Synergus crassicornis S70		100	100	100
	Synergus clandestinus S57		100	64	100
3	*Synergus gallaepomiformis/pallipes* S24	100			
	Synergus gallaepomiformis/pallipes S25				
	Synergus pallidipennis S12				
	Synergus pallicornis/pallipes S22				
	Synergus gallaepomiformis S14				
	Synergus gallaepomiformis S15				
	Synergus gallaepomiformis/pallicornis S16				
	Synergus pallipes S82				

MOTU number	Morphological species identity	Cut-off (bp)			
		COI	*cytb*		
		41bp	22bp	19bp	8bp
3 cont.	*Synergus gallaepomiformis/pallicornis* S19				
	Synergus gallaepomiformis/pallicornis S18				
	Synergus gallaepomiformis/pallicornis S17				100
	Synergus pallicornis S20			36	100
	Synergus pallipes S87				
	Synergus pallipes S27		55	100	46
4	*Synergus umbraculus* S1	100			
	Synergus umbraculus S2				
	Synergus umbraculus S100				
	Synergus umbraculus S102				
	Synergus sp. S109				
	Synergus umbraculus S197				
	Synergus umbraculus S103				
	Synergus hayneanus S81				
	Synergus umbraculus S101				
	Synergus flavipes S38				
	Synergus pallidipennis S181				24
	Synergus umbraculus S201				
	Synergus umbraculus S5				
	Synergus umbraculus S198				69
	Synergus umbraculus S199				
	Synergus umbraculus S196				
	Synergus umbraculus #			11	100
	Synergus umbraculus S3		31	100	100
	Synergus gallaepomiformis S169				
6	*Synergus consobrinus* S56	100	45	1	100
5	*Synergus hayneanus* S9	100			
	Synergus hayneanus/reinhardi S6				
	Synergus hayneanus/reinhardi S7				
	Synergus hayneanus/reinhardi S75				
	Synergus hayneanus/reinhardi S76				
	Synergus hayneanus S77				
	Synergus reinhardi S182				
	Synergus reinhardi S184				
	Synergus reinhardi S185				
	Synergus reinhardi S186				
	Synergus reinhardi S187				
	Synergus reinhardi S189				
	Synergus reinhardi S190				
	Synergus reinhardi S191				
	Synergus reinhardi S192				
	Synergus reinhardi S193				
	Synergus reinhardi S194				
	Synergus reinhardi S188				100
7	*Synergus hayneanus* S78	100	55	99	100
	Synergus hayneanus S79				
	Synergus hayneanus S80				
	Synergus hayneanus S172				
	Synergus hayneanus S173				
	Synergus hayneanus S174				
	Synergus hayneanus S175				
	Synergus umbraculus S146				
	Synergus umbraculus S147				
	Synergus umbraculus S148				
	Synergus umbraculus S149				
	Synergus umbraculus S150				
	Synergus umbraculus S151				
8	*Synergus variabilis* S99	100	100	100	100
	Synergus flavipes S39				
	Synergus flavipes S131				
	Synergus variabilis S195				
	Synergus acsi S112				100

MOTU number	Morphological species identity	Cut-off (bp)			
		COI	*cytb*		
		41bp	22bp	19bp	8bp
9	*Synergus hayneanus* S8	100	100	100	100
	Synergus hayneanus/umbraculus S10				
	Synergus hayneanus/umbraculus S11				
	Synergus reinhardi S183				
10	*Synergus physocerus* S28	7	100	100	100
	Synergus physocerus S29				
	Synergus physocerus S60				
	Synergus physocerus S61				
	Synergus pallipes S87	93			
11	*Synergus diaphanus* S30	100	100	100	100
	Synergus diaphanus S31				
	Synergus diaphanus S167				
	Synergus pallidipennis S13				
12	*Synergus thaumacerus* S53	100	100	100	100
	Synergus thaumacerus S54				
13	*Saphonecrus lusitanicus* S66	72	52	100	100
	Saphonecrus lusitanicus S67				
	Saphonecrus lusitanicus S68	28	48	100	100
14	*Synergus mikoi* S105	100	100	100	100
	Synergus mikoi S106				
16	*Synergus pallipes* S62	100	100	100	100
	Synergus pallipes S63				
17	*Saphonecrus undulatus* S46	100	100	100	100
	Saphonecrus undulatus S47				100
18	*Synergus plagiotrochi* S65	100	100	100	100
19	*Synergus apicalis/tibialis* S40	100	100	100	100
	Synergus apicalis/tibialis S51				
20	*Synergus japonicus* S92	100	100	100	100
	Synergus japonicus S96				
21	*Synergus chinensis* S90	100	100	100	100
22	*Synergus xiaolongmeni* S94	100	100	100	100
23	*Synergus incrassatus* S59	100	100	100	100
24	*Saphonecrus connatus* S50	100	100	100	100
25	*Saphonecrus haimi* S49	100	100	100	100
26	*Synergus gallaepomiformis* #	100	100	100	100
	Synergus umbraculus S200				
	Synergus gallaepomiformis S168				
	Synergus pallicornis S179				
	Synergus gallaepomiformis S170				
	Synergus gallaepomiformis S171				
	Synergus bechtoldae S107				100
	Synergus bechtoldae S108				
	Synergus umbraculus S142				100
	Synergus umbraculus S143				
	Synergus umbraculus S144				
	Synergus umbraculus S145				
	Synergus incrassatus S176				
27	*Synophrus politus* S32	100	56	62	100
	Synophrus politus S205				
	Synophrus politus S207				
	Synophrus politus S208				
	Synophrus politus S209				
	Synophrus politus S206			38	100
	Synophrus politus S210		43	100	100
28	*Synophrus pilulae* S33	100	100	100	100
29	*Synergus pallicornis* S20	-	100	100	100
30	*Synergus pallidipennis* S12	-	100	100	100

Table 8.4 MOTU clustering for *coxI* (41 bp cut-off) and 28SD2 (1 bp cut-off) sequences from the set of samples for which sequences were available for all genes in this study. Each MOTU is separated by a solid line. The percentage of 100 random resampling orders for which the identified sequences were placed in the same MOTU is presented for each cut-off. Dashed lines indicate MOTU divisions supported by less than 100 percent of resampling orders. In such cases the value above the dashed line is the percentage of resampling orders for which all sequences between solid lines were placed in a single MOTU, the value below the dashed line is the percentage of resampling orders for which the sequences were divided into two separate MOTUs. Samples for which all three gene sequences were available have been shaded.

MOTU Number	Morphological species identity	Cut-off (bp)	
		COI 43bp	28SD2 1bp
0.1	*Ceropteres cerri* S36	100	100
	Ceropteres cerri S37		
0.2	*Ceroptres clavicornis* S34	100	100
	Ceroptres clavicornis S35		
0.4	*Rhoophilus loewi* S164	100	100
	Rhoophilus loewi S165		
1	*Synergus apicalis/tibialis* S48	100	45
	Synergus apicalis/tibialis S52		
	Synergus umbraculus S139		
	Synergus umbraculus S140		
	Synergus umbraculus S141		
2	*Synergus clandestinus* S57	100	
	Synergus clandestinus S58		
	Synergus crassicornis S70		
	Synergus crassicornis S69		31
3	*Synergus gallaepomiformis/pallipes* S24	100	89
	Synergus gallaepomiformis/pallipes S25		
	Synergus pallidipennis S12		
	Synergus pallicornis/pallipes S22		
	Synergus gallaepomiformis/pallicornis S19		
	Synergus gallaepomiformis/pallicornis S18		
	Synergus gallaepomiformis/pallicornis S17		
	Synergus pallipes S27		
	Synergus pallicornis S20		
	Synergus pallicornis/pallipes S23		
	Synergus pallipes S26		
	Synergus gallaepomiformis/pallicornis S16		
	Synergus gallaepomiformis S15		
	Synergus gallaepomiformis S14		
	Synergus pallicornis S21		
	Synergus apicalis/tibialis S41		
16	*Synergus pallipes* S62	100	11
	Synergus pallipes S63		
4	*Synergus umbraculus* S100	100	74
	Synergus umbraculus S102		
	Synergus sp. S109		
	Synergus umbraculus S104		
	Synergus umbraculus S1		11
	Synergus flavipes S38		
	Synergus umbraculus S2		
	Synergus umbraculus S3		
	Synergus umbraculus S152		
	Synergus umbraculus S154		
	Synergus umbraculus S155		
	Synergus umbraculus S156		
	Synergus umbraculus S157		
	Synergus umbraculus S158		
	Synergus umbraculus S159		
	Synergus umbraculus S160		
	Synergus umbraculus S161		

MOTU Number	Morphological species identity	Cut-off (bp)	
		COI 43bp	28SD2 1bp
5	*Synergus hayneanus* S9	100	100
	Synergus hayneanus/reinhardi S6		
	Synergus hayneanus/reinhardi S7		
6	*Synergus consobrinus* S56	100	96
	Synergus consobrinus S55		
	Synergus thaumacerus S42		
	Synergus thaumacerus S43		
	Synergus pallipes S83		
	Synergus consobrinus S44		
	Synergus consobrinus S45		
7	*Synergus hayneanus* S78	100	100
	Synergus hayneanus S79		
	Synergus hayneanus S80		
	Synergus umbraculus S146		
	Synergus umbraculus S149		
	Synergus umbraculus S162		
8	*Synergus variabilis* S99	100	100
	Synergus flavipes S39		
	Synergus flavipes S131		
	Synergus acsi S112		
	Synergus variabilis S98		
9	*Synergus hayneanus* S8	100	100
	Synergus hayneanus/umbraculus S10		
	Synergus hayneanus/umbraculus S11		
10	*Synergus physocerus* S28	100	100
	Synergus physocerus S29		
	Synergus physocerus S60		
	Synergus physocerus S61		
11	*Synergus diaphanus* S30	100	100
	Synergus diaphanus S31		
	Synergus pallidipennis S13		
12	*Synergus thaumacerus* S53	100	100
	Synergus thaumacerus S54		
	Synergus thaumacerus S138		
13	*Saphonecrus lusitanicus* S66	72	100
	Saphonecrus lusitanicus S67		
	Saphonecrus lusitanicus S68	28	
14	*Synergus mikoi* S105	100	100
31	*Synophrus politus* S135	-	100
17	*Saphonecrus undulatus* S46	100	100
	Saphonecrus undulatus S47		
18	*Synergus plagiotrochi* S65	100	100
	Synergus plagiotrochi S64		
19	*Synergus apicalis/tibialis* S40	100	100
20	*Synergus japonicus* S92	100	100
	Synergus japonicus S96		
	Synergus japonicus S97		
21	*Synergus chinensis* S90	100	100
22	*Synergus xiaolongmeni* S94	100	100
23	*Synergus incrassatus* S59	100	100
24	*Saphonecrus connatus* S50	100	100
25	*Saphonecrus haimi* S49	100	100
26	*Synergus gallaepomiformis* #	100	100
27	*Synophrus politus* S32	100	100
28	*Synophrus pilulae* S33	100	100
31	*Synergus irani* S113	-	100
	Synergus irani S114		
	Synergus irani S115		
32	*Synergus umbraculus* S143	-	39
	Synergus umbraculus S145		
	Synergus bechtoldae S107		61

8.3.2 Phylogenetic reconstruction

8.3.2.1 Phylogenetic utility

Of the three molecular markers used in phylogeny reconstruction, all had fewer than 10% of quartets in the unresolved central portion of the likelihood map (Figure 8.1). *coxI* had the highest phylogenetic utility with 96.5% of quartets in the well-resolved regions towards the corners, followed by *cytb* (92.2%) and lastly 28SD2 (86.8%).

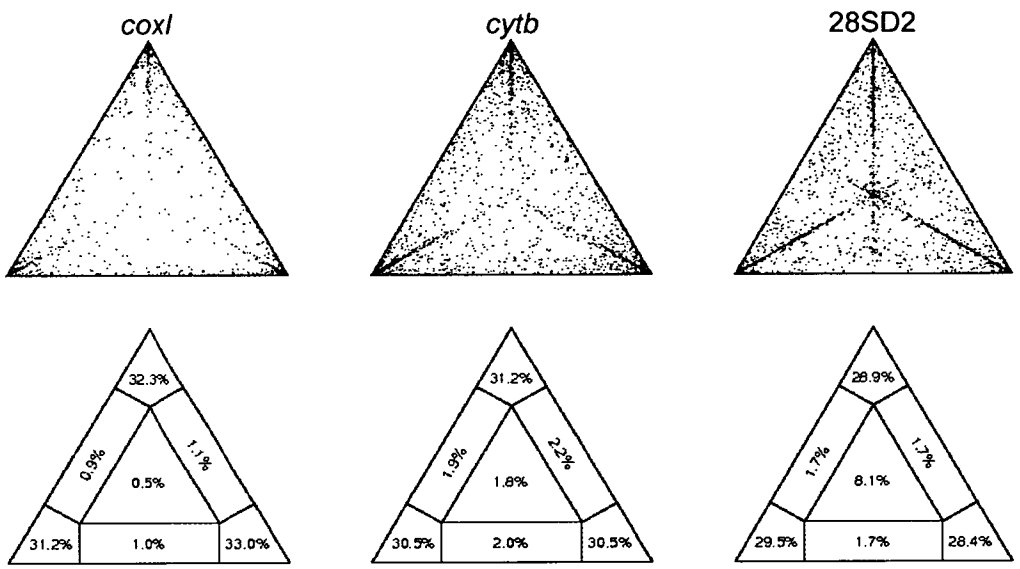


Figure 8.1 Likelihood maps produced using Tree Puzzle 5.0 (Schmidt *et al.* 2002) under the HKY model of evolution showing the phylogenetic utility of the three molecular markers used in this study. The upper row shows the distribution of likelihoods for each of the possible quartets of taxa for each gene. The lower row indicates the proportion of quartets that were poorly resolved (central portion) and well resolved (corners) for each gene.

8.3.2.2 Congruence of phylogenetic signal in *coxI*, *cytb* and 28SD2 sequences

Marginal likelihoods of three-gene phylogenies constructed with shared topology across partitions and with partition topologies unlinked are summarised in Table 8.5. All models other than the independent topology for each gene model are rejected so the three genes each support different topologies. Of the three genes, *coxI* is most consistent with the three gene phylogeny, providing at least 70% support for nodes at

all depths (Figure 8.2). Support declines with depth for both of the mitochondrial genes before increasing to 100% support for the basal split between the Synergus complex and *Rhoophilus*, which is resolved by all three genes. The level of support increases with depth for 28S D2.

Table 8.5 Test for topological congruence between the three genes. Bayes Factor scores >20 indicates strong evidence so there is conflict between each partition and it appears that the greatest conflict is between cyt*b* and *coxI*. For each analysis, | indicates that topology is unlinked between partitions, - indicates that topology is linked.

Partitions	Harmonic mean log-likelihood	Bayes Factor comparison with fully unlinked model
cyt*b* | *coxI* | 28SD2	-14221.37	0
cyt*b* - *coxI* | 28SD2	-14718.35	-993.96
cyt*b* | *coxI* - 28SD2	-14272.51	-102.28
cyt*b* - 28SD2 | *coxI*	-14464.47	-486.2
cyt*b* - *coxI* - 28SD2	-14798.16	-1153.58

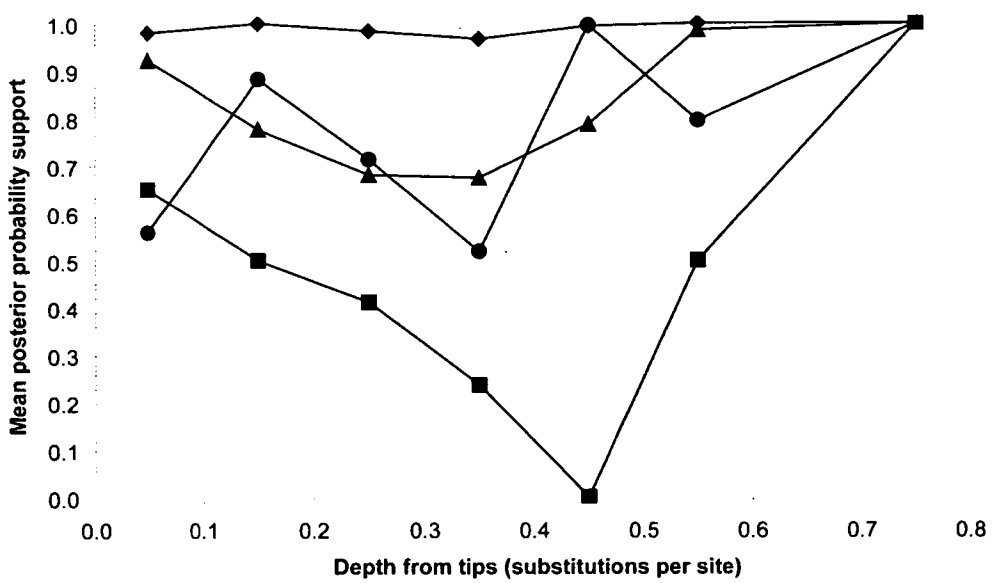


Figure 8.2 The relationship, for each of the single gene datasets (cytb, squares; coxI, triangles; and 28S D2, circles), between evolutionary depth and posterior probability support for nodes resolved in the three-gene phylogeny (diamonds).

8.3.2.3 The influence of missing data on phylogeny reconstruction

The marginal likelihood of the phylogenetic hypothesis of relationships among the Synergini declines with the inclusion of an increasing percentage of missing data in the supermatrix (Figure 8.3). The extent of the reduction in the marginal likelihood was also affected by the identity of taxa for which gene sequences were removed. Since no readily interpretable pattern could be identified to explain this variation within sets of phylogenies with the same amount of missing data, a conservative approach was taken to the assessment of the amount of missing data that could be included in an incomplete supermatrix without affecting the phylogenetic hypothesis.

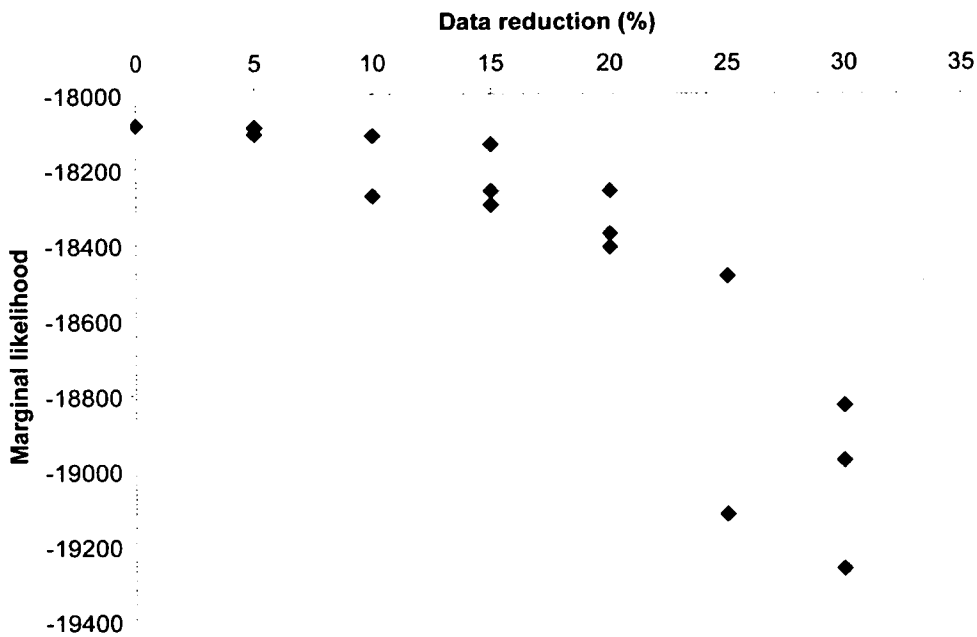


Figure 8.3 The effect of missing data on the marginal likelihood of Bayesian phylogenetic inference. For each level of data reduction, three random sets of sequence data were removed from the complete supermatrix before phylogenetic inference. Marginal likelihoods of the resulting phylogenetic hypotheses were then calculated as the harmonic mean of the individual likelihoods of the sampled phylogenies, calculated using the complete supermatrix data in PAUP*.

When marginal likelihoods were calculated under the complete supermatrix for topologies inferred under the reduced and complete maxdata supermatrices, only the 5% reduced set gave an estimate of marginal likelihood (-18088.6) sufficiently close to that of the phylogeny inferred under the complete maxdata supermatrix (-

18083.4) to be considered acceptable under a relaxed interpretation of Bayes factors (BF = 10.4). Variation between supermatrices was also much lower in the 5% reduced set (σ = 9.85) than in any other group (σ > 75 for all other reduced sets). The reduction in phylogenetic resolution with increased missing data is illustrated in Figure 8.4. As the percentage of missing data is increased above 5%, resolution of nodes inferred under the complete supermatrix is detrimentally affected, initially in mid-range (0.25-0.4 substitutions per site) nodes (at 10%) then at both recent (0.05-0.2 substitutions per site) and mid-range nodes (at 15%). Critically all nodes that are resolved using the complete supermatrix receive at least 75% posterior probability support from the 5% reduced set. Once the proportion of missing data is increased to 10% or greater, many of these nodes receive less than 50% posterior probability support, indicating that different mid-level clades are inferred under these missing data supermatrices. Resolution of basal relationships is relatively unaffected by the inclusion of missing sequence data.

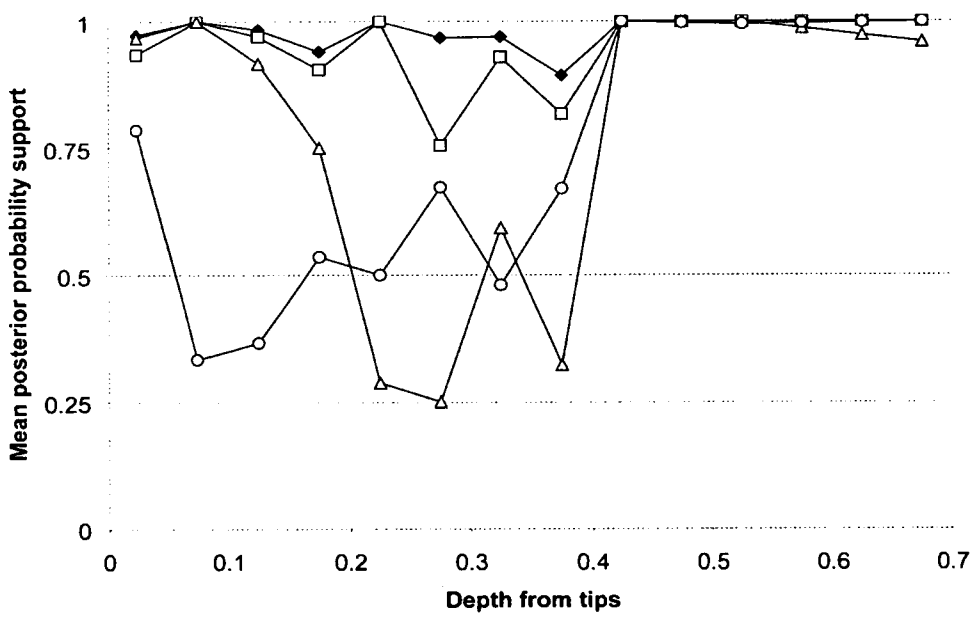


Figure 8.4 The variation of support (posterior probability) for clades resolved by the maxdata supermatrix with evolutionary depth for varying degrees of missing data: complete supermatrix, filled diamonds; 5% missing data, open squares; 10% missing data, open triangles; and 15% missing data open circles. Graphical representation was simplified by averaging support over node depth ranges of 0.05 units.

8.3.2.4 Phylogenetic relationships inferred from the inclusive supermatrix

Incorporation of at least one taxon per MOTU required an inclusive supermatrix with 4.2% missing data. No conflicting relationships were inferred relative to the maxdata supermatrix phylogeny so only the inclusive supermatrix phylogeny is presented (Figure 8.5).

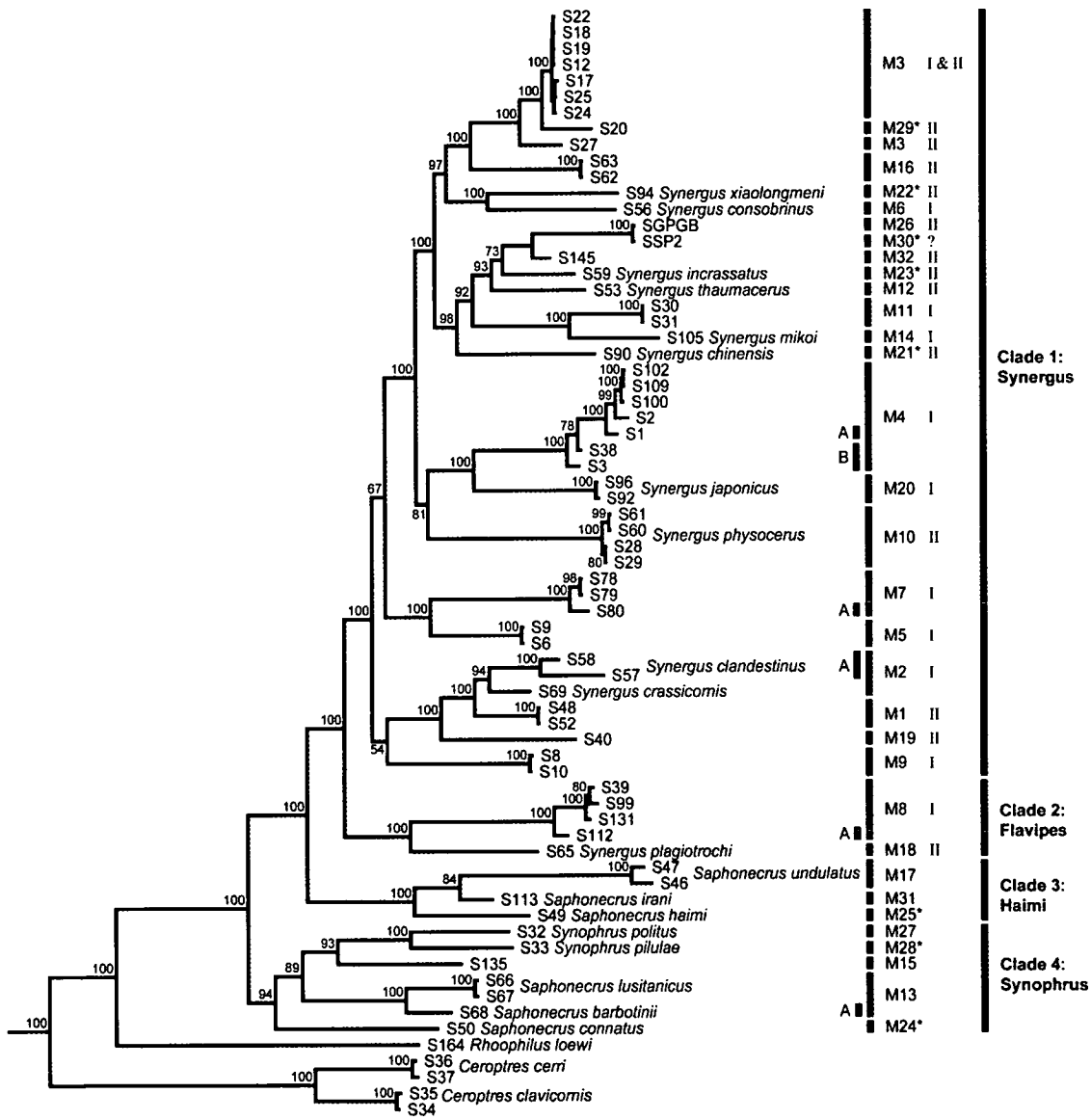


Figure 8.5 Bayesian consensus phylogeny of the 5% missing data three gene (*coxI*, *cytb* and 28SD2) supermatrix, including at least one sequence for each MOTU defined in Tables 8.3 to 8.5. Node support values are posterior probabilities (expressed as percentages). MOTU numbers (M*x*) are for *coxI* at the 41 bp, inclusive, cut-off (Table 8.3). *CoxI* MOTUs defined at lower cut-offs are indicated with letters. Asterisks denote singleton MOTUs. Mayr sections I and II in *Synergus* are labelled for each MOTU.

Of the three genera in the Synergus complex, *Synophrus* and *Synergus* were reconstructed as monophyletic groups, while *Saphonecrus* is divided between two discrete lineages. Four main clades within the Synergus complex are supported by the molecular data with *Rhoophilus* as the sister group. The four clades, labelled in Figure 8.5, are as follows:

1. The **Synophrus clade** contains *Synophrus politus* and three *Saphonecrus* species, *S. barbotini, S. connatus* and *S. lusitanicus.*
2. The **Haimi clade** contains the 3 remaining Western Palaearctic *Saphonecrus* species (*Saphonecrus haimi, S. undulatus* and a recently described new species *S. irani)*. This clade is sister to those containing *Synergus* species.
3. The **Flavipes clade** contains 5 recognised *Synergus* species (*S. plagiotrochi, S. flavipes, S. variabilis, S. acsi, S. palmirae*), and is sister group to the remaining *Synergus* species. On the basis of morphology a newly described oak inquiline genus, *Ufo* (Melika et al., 2005), may well belong to this clade.
4. The **Synergus clade** contains all remaining *Synergus* species.

Placement of species in Mayr's Sections I and II is shown in Figure 8.5. The separation of *Synergus* into morphology-based sections is not supported by data for any gene (data not shown).

8.4 Discussion

8.4.1 Utility and nuclear gene support for *coxI* barcodes

Previous studies have suggested a minimum threshold of 2-3% sequence divergence for designation of separate MOTUs using nuclear SSU data (Blaxter 2004). Appropriate levels of *coxI* divergence to define MOTUs that retained the distinction between species within the genera *Ceroptres* and *Rhoophilus* were between 2.4 and 6.2 % sequence divergence. While this is compatible with the general recommendation, the approach adopted here, of calibration with reliable morphologically identified taxa, allows meaningful comparison with current species status. The importance of calibration rather than acceptance of an arbitrary value is

highlighted by the fact that equivalent cut-offs in *cytb* were 1.8% to 5.4%, so even alternate mitochondrial genes cannot be assumed to have the same threshold. All three genes resolved a broadly congruent set of MOTU clusters, with the 28SD2 set more similar to *coxI* than those resolved using *cytb*. Comparison between MOTU sets defined using *coxI* and *cytb* was possible across the full range of sequence divergences. The RF distance between the MOTU sets increased as the cut-off was reduced, reflecting greater potential for misassignment of sequences at lower thresholds. Although *coxI* and 28SD2 clusters were most similar, it was only possible to select an equivalent 28SD2 cut-off to the 6.2% *coxI* threshold. This was equivalent to 1 bp difference across the entire amplified 28SD2 fragment. The cut-off of 2-3% divergence has been suggested to reduce the impact of base-calling errors in high-throughput barcoding studies (Blaxter 2004). The ability to resolve similar clustering to *coxI* using a single base-pair difference in 28SD2 highlights the increase in sequencing accuracy obtained by sequencing in both directions and checking the sequence traces by eye. Given the congruence between *coxI* and the nuclear marker, it is likely that the *coxI* MOTU clusters are more representative of taxonomic status than the *cytb* sequences. A further argument for using *coxI* over *cytb* is that the CB1/CB2 primers used in this study did not amplify in *ca.* 25% of specimens, suggesting that universal primers for *cytb* may be substantially harder to identify than those widely acknowledged for *coxI* (Hebert *et al.* 2003).

8.4.2 Conflict between groupings based on morphology and sequence data

8.4.2.1 Cryptic variation revealed in morphological species

Many of the morphologically identified species showed high intraspecific *coxI* sequence variation. Sequences from seven of these species were separated into multiple MOTU clusters and several specific examples are discussed in turn below.

The genus *Synophrus* contains inquilines that attack very small host galls (Pujade-Villar *et al.* 2003) and modify them extensively, producing large and very woody galls on section *Cerris* oaks. Three species are recognised in the genus, of

which only *S. politus* is well-known (Stone *et al.* 2002b, Csóka *et al.* 2005). The analysis included *coxI* data for 4 individuals of this species, two of which (S32 and S33) were sampled at the same time from the same host oak (*Q. cerris*) at the same locality (Sopron, Hungary). These two specimens have highly divergent *coxI* sequences, and were assigned to separate MOTU clusters (Clusters 28 and 29) at all cut-offs, suggesting that they may represent two distinct species. These in turn are highly divergent from a third MOTU (Cluster 15), comprising specimens S134 and S135 from Spain and Algeria, respectively. Further sampling is required to assess the extent of within-species diversity within the *Synophrus politus* group, but comparison with levels of intraspecific diversity within the *Synergus* complex as a whole suggests the existence of 3 species.

A total of 19 specimens were identified morphologically as *S. hayneanus* or as showing characters shared between this species and either *S. reinhardi* or *S. umbraculus*. *coxI* data supported allocation of these specimens to 3 MOTU clusters, with evidence for a further possible division at the 2.4% cut-off, which is at the interface between inter- and intra-specific variation: (i) Cluster 5 contains seven specimens, four of which show some morphological attributes of *S. reinhardi*; (ii) Cluster 7 contains 3 specimens, identified as *S. hayneanus* and shows a possible division into two sub-clusters (Cluster 7A); and (iii) Cluster 9 contains 4 specimens, 2 of which show some morphological attributes of *S. umbraculus*. Each of these clusters was supported by the 28SD2 sequence data, however, *cytb* did not differentiate absolutely between Clusters 5 and 7 (or Cluster 6, *Synergus consobrinus*) at the most inclusive cut-off. Clusters 5 and 7 were separated by the *cytb* data at lower cut-offs. Since only one hind leg of each specimen was used for DNA extraction, it was possible to re-examine specimens attributable to each of these clusters. The taxa could not be differentiated using currently recognised morphological criteria. This suggests both that the characters used to differentiate the morphological taxa involved (*S. reinhardi, S. hayneanus*, *S. umbraculus* and possibly *S. consobrinus*) need to be reassessed, and that previously undefined cryptic taxa exist in this group.

Synergus umbraculus is one of the most abundant inquilines in cynipid oak galls, with more than 30 known host oak galls. Specimens attributed to this morphospecies (n=48, Appendix 8) can be divided among three to four clusters on the basis of *coxI*, depending on the cut-off chosen. This suggests that this morphospecies contains cryptic sibling species.

8.4.2.2 Allocation of multiple morphospecies to the same molecular taxon

The analyses revealed multiple cases of different morphospecies being allocated to the same molecular taxon, suggesting phenotypic variability within a single lineage. This is demonstrated clearly by Cluster 3, which contained 20 specimens attributed to five species (*S. nervosus, S. albipes, S. gallaepomiformis, S. pallicornis* and *S. pallidipennis*). With the exception of one specimen (*S. pallipes* S87), which was assigned to a separate MOTU (Cluster 3B) at lower cut-offs, and two further specimens (*S. pallicornis* S21 and *S. apicalis/tibialis* S41) for which there was weak support for assignation to a separate MOTU (Cluster 3A) at the lowest cut-off level, these specimens all had similar or identical *coxI* sequences. The 28SD2 sequence data did not provide strong support for separation of Cluster 3 from Cluster 16, which contains further specimens identified as *S. pallipes*. The possession of identical sequences by morphologically diverse specimens may be attributable to the fact that inquiline cynipids with multivoltine lifecycles (as for these species) show substantial phenotypic variation within and between generations. Some variation in colour is commonly observed from variation in temperature or humidity (Nieves-Aldrey 1986, Pujade-Villar 1992c, Wiebes-Rijks 1979), and within-generation variation may be caused by differences in quantity and/or quality of larval food. Characters used to describe the taxa compounded in Cluster 3 may have been inadequate to separate variation within taxa (assuming they are genuinely different) from divergence between taxa. Alternatively, the specimens attributable to Cluster 3 may simply represent redescriptions of phenotypic plasticity in a single taxon.

If the morphologically identified specimens in Cluster 3 do in fact represent sibling species, the lower than expected within-cluster diversity for COX1 could be attributed to a wide range of demographic effects, including the relatively enhanced impact of a population bottleneck on the lower effective population size of a mitochondrial marker (1/4 that of a nuclear gene), or a selective sweep of mitochondrial genetic diversity (Hurst & Jiggins 2005). Such sweeps can be associated with infection by *Wolbachia* and other maternally inherited symbionts (Hurst & Jiggins 2005), and six *Synergus* species, including *S. gallaepomiformis* in Cluster 3, are known to be infected with *Wolbachia* (Rokas *et al.*, 2002a; the phenotypic impact of *Wolbachia* in *Synergus* remains unknown). The *cytb* data, however, reveals distinct MOTUs within this clade, inferring that the low sequence diversity is not found across the entire mitochondrial genome.

Two members of Cluster 8, *Synergus variabilis* and *S. flavipes*, can be clearly differentiated morphologically (Pujade-Villar *et al.* 2003). One specimen of *S. flavipes* (S38) was placed in a separate MOTU (Cluster 4). The remaining sampled specimens of these morphological species showed low *coxI* sequence divergence, however, and have been placed in a single MOTU cluster despite their geographically distant origins (Hungary and Iran). This molecular grouping is supported by both 28SD2 and *cytb* data. Failure to discriminate *S. variabilis* and *S. flavipes* with COX1 barcodes reinforces the conclusion from Cluster 3 that morphological diversity can be a poor indicator of species in this group. Two other members of the *S. flavipes* clade, *S. acsi* and *S. palmirae*, are newly described morphological species from Iran (Sadeghi *et al.* 2006). The *coxI* sequence divergence between these taxa based on very limited sampling does not support separate species status for these morphological taxa. It is possible that *S. flavipes/S. variabilis/S. acsi/S. palmirae* may represent a single species as they were only differentiated at the lowest cut-off with each of the mitochondrial genes, with *S. acsi* and *S. palmirae* assigned to Cluster 8A. However, the morphological distinctness of Cluster 8A is remarkable. Closely related but distinct biological entities can share

very similar *coxI* sequences (e.g. Johnson & Cicero 2002), and further study of this lineage is required before biological species boundaries can be inferred.

8.4.2.3 Identification errors

The females of *S. consobrinus* and *S. thaumacerus* are very similar, and current morphology-based keys are known to have difficulty distinguishing them (Pujade-Villar *et al.* 2003). Barcode analysis shows that these two species constitute two separate clusters across the full range of cut-offs for *coxI*, *cytb* and 28SD2. However, two specimens identified as *S. thaumacerus* were placed in the *S. consobrinus* cluster, and are thus likely to represent mistaken identifications. This shows the clear utility of barcodes in allocating morphologically similar specimens to biologically discrete taxa.

8.4.3 Phylogenetic relationships within the *Synergus* complex

8.4.3.1 Phylogenetic utility and congruence of molecular markers

Likelihood mapping demonstrated that all three genes have sufficient phylogenetic signal to resolve a high proportion of quartets so all are suitable for the phylogenetic reconstruction of this group. Resolution was greatest for *coxI*, reflecting the utility of this gene in resolving species level divisions during the barcoding analysis. In contrast 28SD2 was the least resolved of the three markers. 28SD2 accumulates mutations at a much slower rate than the mitochondrial genes so this gene is unable to resolve intraspecific relationships. The advantage of including this gene in phylogenetic reconstruction is that, unlike the mitochondrial genes, whose resolving power decreases with evolutionary depth, the resolution of 28SD2 increases at deeper nodes. The three markers each support different phylogenetic topologies and therefore phylogenetic relationships between the taxa will be discussed in the context of the multigene phylogenies, which may be expected to approximate the species phylogeny.

8.4.3.2 The influence of missing data

Inferences based on each of the three genes differ both at the level of defining species-level MOTU clusters and in the reconstruction of deeper phylogenetic relationships. Ideally, multigene phylogenies should be constructed from a complete supermatrix (Gatesy *et al.* 2004), however, sequence data could not be obtained for all three genes for all of the specimens in this study. Here, the impact of missing data on the reconstruction of a known phylogeny is assessed to determine the acceptable level of missing data for the specific combination of genes and taxonomic group included in this study. All phylogenies reconstructed for the set of species for which sequence data were available for each gene, but with some data deliberately excluded, had an increased marginal likelihood. This reduction in marginal-likelihood was acceptable only up to the 5% missing data level, i.e. 95% non-random data. The detrimental effect of including a greater proportion of missing data prevented analysis of all data as a single supermatrix. Taxa with missing data for one or two genes were incorporated by ensuring that each MOTU was included in the maxtaxa supermatrix. Including at least one taxon per MOTU required the introduction of 4.2% missing data so all MOTUs could be included in a single analysis without detrimentally affecting the accuracy of the phylogenetic reconstruction.

8.4.3.3 Phylogenetic relationships

All phylogenetic hypotheses reconstruct *Synophrus* and *Synergus* as monophyletic, while *Saphonecrus* is polyphyletic. The three *Saphonecrus* species allied to *Synophrus* (*S. barbotini, S. connatus* and *S. lusitanicus*) should probably be transferred to the genus *Synophrus*, while the distinct lineage comprising *Saphonecrus haimi* and *S. undulatus* could retain the genus name *Saphonecrus*. Within the genus *Synergus*, Mayr's long-accepted morphologically based sections are entirely unsupported by molecular data, and should be abandoned.

8.5 Conclusion and prospects

CoxI barcodes are shown to be an effective in allocating inquiline specimens to well-resolved molecular taxa. Morphology-based taxonomy of the *Synergus* group of inquilines is shown to be unsound at the level of putatively fundamental species groups (Mayr's Sections 1 and II) and of species. While some taxa can be identified reliably on the basis of morphology, others clearly cannot, since the specimens in this study were identified by the authors of the current morphological keys. This problem is at least in part due to phenotypic plasticity within inquiline taxa associated with differentiation between and within generations, and such variation must clearly be considered in greater depth when new taxa are described and existing taxa revised. This study revealed both Type I and Type II errors in conventional taxonomic treatments (Quicke, 2004). In Cluster 3, morphology suggested separation into discrete taxa where none exists (Type I error), while cryptic lineages were revealed within single morphological species (Type II error). The latter (as in *S. politus*) suggests that further biological diversity may exist within recognised taxa, and the basis of such diversity (for example, the existence of host-specific sibling taxa, as in *S. umbraculus*) needs to be investigated. Finally, the identification of taxonomist errors (in the case of *S. consobrinus* and *S. thaumacerus*) shows the value of molecular barcodes in a taxonomically challenging group.

Chapter 9

Concluding remarks

This thesis began with three major aims: (i) to utilise phylogenetic approaches to address a specific set of phylogeographic questions; (ii) to develop bioinformatic methods; and (iii) to improve understanding of the evolutionary history of the Western Palaearctic oak gallwasps and oak inquilines. This chapter discusses the progress made towards achieving these aims and highlights directions for future research.

9.1 Answering phylogeographic questions

Understanding of species' phylogeographic histories has proved important in setting conservation priorities through identifying: (i) areas of high genetic diversity; (ii) evolutionary significant units; (iii) cryptic species; and (iv) appropriate populations for reintroduction. However, these important aims have frequently been addressed with limited sampling, which risks excluding areas of high diversity and may adversely affect the overall conclusions of a study. This thesis has presented a review of the phylogeographic literature on species distributed across the Western Palaearctic (Chapter 3), which showed that 21 out of 79 studies did not include sampling from all areas of the species' range.

An important goal of this thesis has been to resolve this problem by establishing a longitudinal perspective in phylogeographic studies. The longitudinal approach encourages sampling across a species' entire range, not only to identify all centres of diversity, but also to provide further information to guide conservation priorities. Once a species' entire range has been sampled, it is possible to: (i) determine the relationships between refugial populations; (ii) identify timescales of range expansion; and ultimately (iii) identify geographic origins.

Only 11 out of 79 studies have provided robust support for the geographic origin of a species distributed across the Western Palaearctic (Chapter 3). An eastern origin was supported by 9 of the 11 studies, representing a diverse range of taxa and work in this thesis has provided further evidence for eastern origin in longitudinal phylogeographic studies of oak gallwasp species (Chapters 4 & 7). Further work is required to establish the general importance of this emerging pattern.

Patterns of shared timescale have yet to emerge from studies of diverse taxa. Dates for the origin of taxa from the 11 studies range from the Holocene to the late Miocene (Chapter 3). Results presented in this thesis for the oak gallwasps suggest a common period of origin across a number of species around 2 mya (Chapters 4 & 7). It is possible that the climatic changes associated with the onset of the Pleistocene may have promoted speciation at this time.

Where shared geographic origins are identified across a number of species, the regions concerned are likely to represent cradles of speciation (Chown & Gaston 2000) and should receive high conservation priority. Identification of common timescales and directions of longitudinal range expansion between species may ultimately allow questions of community development, such as whether they spread as coherent trophic assemblages or assemble by local recruitment, to be addressed. If communities are found to have shared longitudinal histories, then conservation priorities must address the requirements of entire communities, rather than focussing on individual species. If community members have concordant phylogeographic histories, then it will be appropriate to study community evolution in an explicitly phylogenetic framework.

Human activity has the potential to alter species distributions through: (i) deliberate introduction; (ii) non-deliberate introduction; and (iii) habitat alteration, creating favourable conditions for unassisted range expansion. Both deliberate introduction (direct trade in the galls of *Andricus kollari*) and habitat alteration (introduction of section *Cerris* oaks to northern Europe) were investigated in this

thesis. Although at least 11 species of oak gallwasp have colonised northern Europe, exploiting the increased range of section *Cerris* oaks over the last 500 years, their longitudinal phylogeographic signatures have remained unaffected (Chapter 7). Direct trade in the galls of *A. kollari*, however, has had a dramatic effect on longitudinal phylogeography (Chapter 5). Phylogeographic analysis revealed a diverse clade in the UK and northern coastal Europe, with genetic diversity that was not sampled from anywhere in the native range, raising the possibility that an entire phylogeographic clade has been traded from its original location, leaving no trace of the source population.

One important consideration in applying longitudinal phylogeography is whether the current distribution can be used to reconstruct a species' history beyond events during the most recent glacial cycle. Human-mediated habitat alteration did not have a significant impact on the longitudinal phylogeography of the oak gallwasps. This suggests that while changes in the area of available habitat (whether through human activity or climatic fluctuations) are important determinants of latitudinal phylogeographic distributions, longitudinal patterns can remain robust to such changes. However, direct trade in the galls of *A. kollari* created a longitudinal phylogeographic distribution that could only be interpreted with knowledge of the scale of historic trade in the species. Further longitudinal phylogeographic studies of species that have been redistributed by humans within and outside of their native ranges, both deliberately and non-deliberately, will help to reveal the extent to which introductions on similar and smaller scales affect species phylogeographic distributions. This will be important in determining whether analogous natural processes may have disrupted longitudinal phylogeography in other species and in understanding the impact of human activity on species conservation.

9.2 Developing bioinformatic methods

Two classes of method have been proposed in this thesis to allow reconstruction of geographic origins and directions of range expansion. Trait mapping techniques were adopted to reconstruct ancestral geographic regions (Chapters 3 & 7). By using

model-based trait mapping, it was possible to implement a model reduction method to infer the most likely directions of range expansion. However, this approach is not suitable for direct comparison of phylogeographic histories between species. Given the potential implications of longitudinal phylogeographic concordance in setting conservation priorities and understanding community development, a second method was proposed to allow direct comparison (Chapter 7). Although this method has only been put forward as a concept, it has the potential to accommodate many of the complexities of phylogeographic comparison, which should allow the method to be applied to a wide range of intraspecific comparisons.

Longitudinal phylogeography requires sampling across a species' entire range and the use of the most closely related outgroup available. In order to meet these requirements, it is important to have an accurate taxonomic and phylogenetic understanding of the species under consideration. Molecular taxonomy and phylogenetics are ideally suited to resolving relationships among morphologically cryptic taxa. However, most methods require data for the same gene(s) for all samples. Molecular taxonomy, in particular, is traditionally based on a single barcoding gene. By developing an approach to assess equivalence in molecular taxonomy across genes (Chapter 8), this thesis has provided a method to apply DNA barcodes in multigene analyses where data are missing for some genes in some taxa.

Model-based phylogenetic reconstruction is, in theory, relatively robust to the inclusion of missing data. Previous studies of the impact of missing data on phylogenetic reconstruction, however, have used alignments of large numbers of genes for relatively small numbers of taxa. This thesis presents an analysis of the impact of missing data on a three-gene dataset for a large number of taxa (Chapter 8). Under these conditions, missing data are found to have a much larger effect on phylogenetic reconstruction. Removal of a similar percentage of data to that in the studies of alignments of large numbers of genes leaves a relatively small overall number of informative characters. This highlights the importance of considering the dimensionality of a sequence alignment in predicting the impact of missing data.

9.3 Improving understanding of the Western Palaearctic oak gallwasps and oak inquilines

In using Western Palaearctic oak gallwasp communities as a model system, this thesis has provided the opportunity to address several specific aspects of oak gallwasp and oak inquiline phylogeography.

9.3.1 Cryptic taxa

This thesis has revealed molecular evidence for cryptic lineages in *Andricus coriarius* (Chapter 4), *Andricus grossulariae* (Chapter 7) and *Cynips quercusfolii* (Chapter 7). Detailed morphological analysis of the cryptic lineages in *Andricus coriarius* revealed diagnostic morphological characters and the two lineages have been elevated to species status: *Andricus coriariformis* (Melika, Challis & Stone 2007) and *Andricus libani* (Melika, Challis & Stone 2007) (Melika *et al.* 2007). Although the adult wasps belonging to the cryptic lineages in *A. grossulariae* and *C. quercusfolii* have yet to be examined in detail, they are similarly divergent from the main lineage as the cryptic lineages in *A. coriarius*, so it seems likely that they will represent further cryptic species. Each of the cryptic lineages was sampled from areas to the east of Europe. Such cryptic diversity suggests that Turkey and the Middle East may represent a cradle of speciation for the oak gallwasps.

9.3.2 Eastern origins

The importance of areas in the east of the Western Palaearctic in oak gallwasp origins is highlighted by the inference, in this thesis, of eastern origin in *A. coriarius sensu stricto* (Chapter 4), *Andricus lucidus* (Chapter 7) and *C. quercusfolii* (Chapter 7). Together with the inference of eastern origin in a further oak gallwasp, *Andricus quercustozae* (reviewed in Chapter 3), this supports the emerging pattern of eastern origin in a diverse set of Western Palaearctic taxa (Chapter 3).

9.3.3 Origins of *Andricus kollari* in northern Europe

Although the galls of *A. kollari* were historically important in trade, the impact of this trade on phylogeographic distributions was unknown. This thesis has demonstrated that, while much of northern Europe has been colonised by unassisted range expansion, an entire phylogeographic clade of *A. kollari* has been traded from an unknown area to the east of the Mediterranean to the UK (Chapter 5). This traded lineage was also sampled in areas of coastal northern Europe, which have been colonised against the prevailing direction of unassisted range expansion.

9.3.4 Taxonomy and phylogeny of the Synergini

This thesis has demonstrated that the species-level taxonomy of the Synergini is not supported by molecular data (Chapter 8). Molecular taxonomy, which was consistent across both nuclear and mitochondrial markers, identified extensive Type I (multiple species names assigned to a single lineage) and Type II (samples belonging to more than one lineage identified as the same species) taxonomic errors. A widely accepted higher level taxonomic division (separation into Mayr's sections I and II) was also unsupported by the phylogenetic relationships within the Synergini. This study should provide a foundation for future work to reassess the appropriate morphological characters for species-level identification.

References

Abrahamson WG, Hunter MD, Melika G & Price PW (2003) Cynipid gall-wasp communities correlate with oak chemistry. *Journal of Chemical Ecology*, **29**, 208-223.

Abrahamson WG, Melika G, Scrafford R & Csóka G (1998) Gall-inducing insects provide insights into plant systematic relationships. *American Journal of Botany*, **85**, 1159-1165.

Ács Z, Melika G, Pénzes Zs, Pujade-Villar J & Stone GN (2007) The phylogenetic relationships between *Dryocosmus*, *Chilaspis* and allied genera of oak gallwasps (Hymenoptera, Cynipidae: Cynipini). *Systematic Entomology*, in press.

Aebi A, Schönrogge K, Melika G, Alma A, Bosio G, Quacchia A. Picciau L, Abe Y, Moriya S, Yara K, Seljak G & Stone GN (2006) Parasitoid recruitment to the globally invasive chestnut gall wasp *Dryocosmus kuriphilus*. Pp 103-122 In: Ozaki K, Yukawa J, Ohgushi T & Price PW (eds.) *Galling arthropods and their associates; ecology and evolution*. Springer, Tokyo.

Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716-723.

Aleixo A (2006) Historical diversification of floodplain forest specialists in the Amazon: a case study with two species of the avian genus *Xiphorhynchus* (Aves: Dendrocolaptidae) *Biological Journal of the Linnean Society*, **89**, 383-395.

Alfaro ME, Zoller S & Lutzoni F (2003) Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Molecular Biology and Evolution*, **20**, 255-266.

Allan GJ, Francisco-Ortega J, Santos-Guerra A, Boerner E & Zimmer EA (2004) Molecular phylogenetic evidence for the geographic origin and classification of Canary Island Lotus (Fabaceae : Loteae). *Molecular Phylogenetics and Evolution*, **32**, 123-138.

Amedegnato C, Chapco W & Litzenberger G (2003) Out of South America? Additional evidence for a southern origin of melanopline grasshoppers. *Molecular Phylogenetics and Evolution*, **29**, 115-119.

Armbruster P, Bradshaw WE & Holzapfel CM (1998) Effects of postglacial range expansion on allozyme and quantitative genetic variation of the pitcher plant mosquito, *Wyeomyia smithii*. *Evolution*, **52**, 1697-1704.

Askew RR (1961) On the biology of the inhabitants of oak galls of Cynipidae (Hymenoptera) in Britain. *Transactions of the Society for British Entomology*, **14**, 237-268.

Askew RR (1965) The biology of the British species of the genus *Torymus* Dalman associated with galls of Cynipidae on oak, with special reference to alternation of forms. *Transactions of the Society for British Entomology*, **16**, 217-232.

Askew RR (1984) The biology of gall wasps, pp. 223–271. *In* Ananthakrischan, T.N. (ed). Biology of Gall Insects. Oxford and IBH Publishing Co., New Delhi, Bombay, Calcutta. 362 pp.

Askew RR (1999) Confirmation of an association of *Synergus* Hartig and *Saphonecrus* Dalla Torre & Kieffer (Hym., Cynipidae) with oak galls of Cecidomyiidae (Diptera). *Entomologist's Monthly Magazine*, **135**, 89-90.

Atkinson RJ (2000) The genetic analysis of natural history, reproductive strategy and population structure in European oak gall wasps (Hymenoptera: Cynipidae). Ph.D. thesis, University of Oxford.

Atkinson RJ, Brown GS & Stone GN (2003). Skewed sex ratios and multiple founding in galls of the oak apple gallwasp Biorhiza pallida (Hymenoptera: Cynipidae). *Ecological Entomology*, **28**, 14-24.

Atkinson RJ, McVean GAT & Stone GN (2002) Use of population genetic data to infer oviposition behaviour: species-specific patterns in four oak gallwasps (Hymenoptera: Cynipidae). *Proceedings of the Royal Society, Series B*, **269**, 383-390.

Atkinson RJ, Rokas A & Stone GN (2007). Longitudinal patterns in species richness and genetic diversity in European oaks and oak gallwasps. In: *Phylogeography in southern European refugia* (ed. S. Weiss S). Kluwer Academic Press.

Avise JC (1994) *Molecular markers, natural history and evolution.* Chapman & Hall, New York.

Avise JC (2000) *Phylogeography. The History and Formation of Species*, Harvard University Press, Cambridge, MA.

Baldauf SL & Palmer JD (1990) Evolutionary transfer of the chloroplast *tufA* gene to the nucleus. *Nature*, **344**, 262-265.

Balfourier F, Imbert C & Charmet G (2000) Evidence for phylogeographic structure in *Lolium* species related to the spread of agriculture in Europe. A cpDNA study. *Theoretical and Applied Genetics*, **101**, 131-138.

Ball RM, Freeman S, James FC, Bermingahm E & Avise JC (1988) Phylogeographic population structure of red-winged blackbirds assessed by mitochondrial DNA.

Proceedings of the National Academy of Sciences of the United States of America, **85**, 1558-1562.

Bănărescu P (1991) *Zoogeography of Fresh Waters, 2: Distribution and Dispersal of Fresh Water Animals in North America and Eurasia*. AULA-Verlag, Wiesbaden.

Bandelt HJ & Dress AWM (1992) A canonical decomposition theory for metrics on a finite set. *Advances in Mathematics*, **92**, 47-105.

Bandelt HJ, Forster, P, Sykes BC & Richards MB (1995) Mitochondrial portraits of human population using median networks. *Genetics*, **141**, 743-753.

Bapteste E, Brinkmann H, Lee JA, Moore DV, Sensen CW, Gordon P, Durufle L, Gaasterland T, Lopez P, Muller M & Philippe H (2002) The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba. *Proceedings of the National Academy of Sciences USA*, **99**, 1414-1419.

Barraclough TG, Nee S (2001) Phylogenetics and speciation. *Trends in Ecology and Evolution*, **16**, 391-399.

Barton NH & Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113-148.

Beavis IC (1988) *Insects and other invertebrates in classical antiquity.* University of Exeter Press, Exeter UK.

Belkhir K (1999) Genetix, version 4.0. A windows programme for population genetic analysis. Laboratoire Génome, Populations, interactions UPR 9060 du CNRS, Université Montpellier 2, Place E. Bataillon, 34095 Montpellier Cedex 05, France.

Bellarosa R, Simeone MC, Papini A & Schirone B (2005). Utility of ITS sequence data for phylogenetic reconstruction of Italian *Quercus* spp. *Molecular Phylogenetics and evolution*, **34**, 355-370.

Bennett KD (1986) The rate of spread and population increase of forest trees during the postglacial. *Philosophical Transactions of the Royal Society of London, Series B*, **314**, 523-531.

Bensasson D, Zhang D-X, Hartl DL & Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology and Evolution*, **16**, 314-321.

Berbee ML, Carmean DA & Winka K (2000) Ribosomal DNA and resolution of branching order among the Ascomycota: how many nucleotides are enough? *Molecular Phylogenetics and Evoution*, **17**, 337–344.

Bergmann PJ & Russell AP (2007) Systematics and biogeography of the widespread Neotropical gekkonid genus *Thecadactylus* (Squamata), with the description of a new cryptic species. *Zoological Journal of the Linnean Society*, **149**, 339-370.

Blaxter M (2004) The promise of a DNA taxonomy. *Philosophical transactions of the Royal Society of London B Biological Sciences*, **359**, 669–679.

Blaxter M, Elsworth B & Daub J (2004). DNA taxonomy of a neglected animal phylum: an unexpected diversity of tardigrades. *Proceedings of the Royal Society of London, Series B*, **271**, 189-192.

Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R & Abebe E (2005) Defining operational taxonomic units using DNA barcode data. *Philosophical transactions of the Royal Society of London, Series B*, **360**, 1935-1943.

Bones TJ (1979) Atlas of fossil fruits and seeds from north central Oregon. *OMSI Occasional Papers in Natural Sciences*, **1**, 1-23.

Boursot P, Din W, Anand R, Darviche D, Dod B, VonDeimling F, Talwar GP & Bonhomme F (1996) Origin and radiation of the house mouse: Mitochondrial DNA phylogeny. *Journal of Evoltionary Biology*, **9**, 391-415.

Brett DW (1960) Fossil oak wood from the British Eocene. *Palaeontology*, **9**, 86-92.

Brooks SE & Shorthouse JD (1998) Developmental morphology of stem galls of *Diplolepis nodulosa* (Hymenoptera: Cynipidae) and those modified by the inquiline *Periclistus pirata* (Hymenoptera: Cynipidae) on *Rosa blanda* (Rosaceae). *Canadian Journal of Botany*, **76**, 365-381.

Brower AVZ (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences USA*, **91**, 6491-6495.

Browicz K & Menitsky GL (1971) Fagaceae. In *Flora des Iranischen hochlandes und der umrahmenden gebirge*, vol. 77 (ed. K. H. Rechinger). Graz, Austria: Akademische Druck-u. Verlagsanstalt.

Brown JH & Lomolino MV (1998) *Biogeography*. Sinauer Associates, Sunderland, Massachusetts.

Broyles SB (1998) Postglacial migration and the loss of allozyme variation in northern populations of *Asclepias exaltata* (Asclepiadaceae). *American Journal of Botany*, **85**, 1091-1097.

Bryant, D & Moulton V (2004) NeighborNet: and agglomerative algorithm for the construction of planar phylogenetic networks. *Molecular Biology and Evolution*, **21**, 255-265.

Camus A (1934-54) *Les Chenes.* 3 volumes. *Monographie du genre Quercus*. Vol. VI Encyclopedie Economique de Silviculture, volumes 6-8. Lechevalier, Paris, France.

Cavalli-Sforza LL & Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics,* **19**,233-257.

Challis RJ, Stone GN, Mutun S, Nieves-Aldrey J-L, Preuss S, Rokas A, Aebi A, Sadeghi E & Tavakoli M (2007) Longitudinal range expansion and cryptic eastern species in the Western Palaearctic oak gallwasp *Andricus coriarius*, *Molecular Ecology*, **16**, 2103-2114.

Chapella IH & Garbellotto M (2004) Phylogeography and evolution in matsutake and close allies inferred by analyses of ITS sequences and AFLPs. *Mycologia*, **96**, 730-741.

Chodjai M (1980) L'étude des Hymenoptères cynipides et les Espèces Cécidogenes dans la Faune des Forêts du Chêne en Iran. *Journal of the Entomological Society of Iran,* Supplement **3**, 1-67.

Chown SL & Gaston KJ (2000) Areas, cradles and museums: the latitudinal gradient in species richness. *Trends in Ecology and Evolution*, **15**, 311-315.

Çiplak B (2003) Distribution of Tettigoniinae (Orthoptera, Tettigoniidae) bush crickets in Turkey: the importance of the Anatolian Taurus mountains in biodiversity and implications for conservation. *Biodiversity and Conservation*, **12**, 47-64.

Çiplak B (2004) Biogeography of Anatolia: the marker group Orthoptera. *Memorie Soc. Entomol. Ital.*, **82**, 357-372.

Claridge MF (1962) *Andricus quercuscalicis* (Burgsdorf) in Britain (Hym., Cynipidae). *The Entomologist*, **95**, 60-61.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657-1660.

Collin R & Cipriani R (2003) Dollo's law and the re-evolution of shell coiling. *Proceedings of the Royal Society of London, Series B*, **270**, 2551-2555.

Comps B, Gömöry D, Letouzey J, Thiébault B & Petit RJ (2001) Diverging trends between heterozygosity and allelic richness during postglacial colonisation in the European beech. *Genetics*, **157**, 389-397.

Cook JM, Rokas A, Pagel M & Stone GN (2002) Evolutionary shifts between host oak species and host plant organs in *Andricus* gall wasps. *Evolution*, **56**, 1821–1830.

Cooper SJB, Ibrahim KM & Hewitt GM (1995) Post-glacial expansion and genome subdivision in the European grasshopper *Chorthippus parallelus*. *Molecular Ecology*, **4**, 49-60.

Cornell HV (1985) Local and regional richness of cynipine gallwasps on California oaks. *Ecology*, **66**,1247-1260.

Crandall KA, Templeton AR (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics*, **134**, 959-969.

Crawley MJ (1997) *Aphelonyx cerricola* Giraud (Hym., Cynipidae) an alien gall-former new to Britain. *Entomologist's Monthly Magazine*, **133**, 61.

Crepet WL & Nixon KC (1985) A preliminary analysis of the fossil history of *Quercus* and other genera of the Fagaceae. *American Journal of Botany*, **72**, 930-931.

Crozier RH & Crozier YC (1993) The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics*, **133**, 97-117.

Csóka G, Stone GN, Atkinson RJ & Schönrogge K (1998) The population genetics of postglacial invasions of northern Europe by cynipid gallwasps (Hymenoptera: Cynipidae). Pp. 280-294 In: Csóka G, Mattson W, Stone GN & Price PW (eds.) *The Biology of Gall-inducing Arthropods*. USDA, St. Paul, Minnesota, USA.

Csóka Gy, Stone GN & Melika G (2005) Biology, ecology, and evolution of gall-inducing Cynipidae. In: Raman A, Schaefer CW & Withers TM (eds.) *Biology, Ecology, and Evolution of Gall-inducing Arthropods*. Science Publishers, Inc., New Hampshire, USA.

Cucchi T, Vigne J-D & Auffray J-C (2005) First occurrence of the house mouse (*Mus musculus domesticus* Schwarz & Schwarz, 1943) in the Western Mediterranean: a zooarchaeological revision of subfossil occurrences. *Biological Journal of the Linnean Society*, **84**, 429-445.

Culling MA, Janko K, Boron A, Vasil'Ev VP, Côté IM & Hewitt GM (2006) European colonization by the spined loach (*Cobitis taenia*) from Ponto-Caspian refugia based on mitochondrial DNA variation. *Molecular Ecology*, **15**, 173-190.

Czerepanov SK (1995) *Vascular plants of Russia and adjacent States (the former USSR)*. Cambridge University Press, Cambridge.

Daghlian CP & Crepet WL (1983) Oak catkins, leaves and fruit from the Oligocene Catahoula formation and their evolutionary significance. *American Journal of Botany*, **70**, 639-649.

Dalla Torre KW & Kieffer JJ (1910) *Cynipidae*. Friedlander & Sohn, Berlin.

Darwin C (1859) *On the origin of species*. John Murray, London.

Dauphin P & Aniotsbeherre J-C (1997) *Les galles de France. Memoires de la Société Linnéene de Bordeaux*, **2**, 1-316.

Dauphin P (1996) Présence en Gironde de la galle sauteuse de Neuroterus saliens (Kollar) (Hymenoptera Cynipidae). *Bulletin de la Société Linnéenne de Bordeaux*, **24**, 181-183.

Dauphin P (1999) Présence en Gironde d'Andricus lucidus (Hartig) (Hymenoptera Cynipidae). *Bulletin de la Société Linnéenne de Bordeaux*, **27**, 18.

Davies N, Villablanca FX & Roderick GK (1999) Bioinvasions of the medfly *Ceratitis capitata*: source estimation using DNA sequences at multiple intron loci. *Genetics*, **153**, 351-360.

Davis PH (1971) Distribution patterns in Anatolia with particular reference to endemism. Pp 15-27 In: Davis PH, Harper PC, Hedge IC (eds.) *Plant Life of Southwest Asia*. Botanical Society of Edinburgh.

Davis PH (ed.) (1965-1985) *Flora of Turkey and the east Aegean islands*. Vol.1-9, Edinburgh University Press, Edinburgh.

Davison A, Birks JDS, Brookes RC, Messenger JE & Griffiths HI (2001) Mitochondrial phylogeography and population history of pine martens *Martes martes* compared with polecats *Mustela putorius*. *Molecular Ecology*, **10**, 2479-2488.

de Lattin G (1967) *Grundriß der Zoogeographie*. Verlag Gustav Fischer, Jena.

De Meeus T & Renaud F (2002) Parasites within the new phylogeny of eukaryotes. *Trends in Parasitology*, **18**, 247-251.

Demesure B, Comps B & Petit RJ (1996) Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. *Evolution*, **50**, 2515-2520.

DeSalle R, Egan MG & Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA bar-coding. *Philosophical transactions of the Royal Society of London, Series B*, **360**, 1905-1916.

Donoghue MJ & Smith SA (2004) Patterns in the assembly of temperate forests around the Northern Hemisphere. *Philosophical transactions of the Royal Society of London, Series B*, **359**, 1633-1644.

Dowton M & Austin AD (1995) Increased genetic diversity in mitochondrial genes is correlated with the evolution of parasitism in the Hymenoptera. *Journal of Molecular Evolution*, **41**, 958–965.

Dreger-Jauffret F, Shorthouse J D (1992) Diversity of gall inducing insects and their galls. Pp 8–33 In: Shorthouse JD, Rohfritsch O (eds.) *Biology of Insect-induced Galls*. Oxford University Press, Oxford, UK.

Dress AWM & Huson DH (2004) Constructing splits graphs. *IEEE Transactions on Computational Biology and Bioinformatics,* **1**, 109-115.

Rambaut A & Drummond AJ (2004) Tracer v1.3, Available from http://evolve.zoo.ox.ac.uk/software.html

Drummond AJ, Ho SYW, Phillips MJ & Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *Public Library of Science Biology*, **4**, e88.

Dumolin-Lapègue S, Demesure B, Fineschi S, Le Corre V & Petit RJ (1997) Phylogeographic structure of White oaks throughout the European continent. *Genetics*, **146**, 1475-1487.

Durand JD, Unlu E, Doadrio I, Pipoyan S & Templeton AR (2000) Origin, radiation, dispersion and allopatric hybridization in the chub *Leuciscus cephalus*. *Proceedings of the Royal Society of London Series B – Biological Sciences*, **267**, 1687-1697.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792-1797.

Edwards A & Cavalli-Sforza L (1964) Reconstruction of evolutionary trees. In: V Heywood & J McNeill (eds.) *Phenetic and Phylogenetic Classification*, pp. 67-76. Publication no. 6. Systematics Association. New York, NY.

Edwards AWF (1972) *Likelihood: an account of the statistical concept of likelihood and its application to scientific inference*. Johns Hopkins Univ. Press, London.

Efron B (1979) Bootstrap methods: another look at the jackknife. *Annals of Statistics*, **7**, 1–26.

Emerson B, Paradis E, Thébaud C (2001) Revealing the demographic histories of species using DNA sequences. *Trends in Ecology and Evolution*, **16**, 707-716.

Englbrecht CC, Freyhof J, Nolte A, Rassmann K, Schliewen U & Tautz D (2000) Phylogeography of the bullhead *Cottus gobio* (Pisces : Teleostei : Cottidae) suggests a pre-Pleistocene origin of the major central European populations. *Molecular Eology*, **9**, 709-722.

Evans D (1967) The bisexual and agamic generations of *Besbicus mirabilis* (Hymenoptera: Cynipidae), and their associate insects. *The Canadian Entomologist*, **99**, 187-196.

Evans D (1972) Alternate generations of gall cynipids (Hymenoptera: Cynipidae) on Garry oak. *The Canadian Entomologist*, **104**, 1805-1818.

Farris JS (1999) Likelihood and inconsistency. *Cladistics*, **15**, 199-204.

Felsenstein J & Kishino H (1993) Is there something wrong with the bootstrap? A reply to Hillis and Bull. *Systematic Biology*, **42**, 193–200.

Felsenstein J (1978a) The number of evolutionary trees. *Systematic Zoology*, **27**, 27-33.

Felsenstein J (1978b) Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology*, **27**, 401-410.

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution*, **17**, 368-376.

Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, **39**, 783-791.

Fernandes GW, Fagundes M, Woodman RL & Price PW (1999) An effect of three-trophic level interactions: plant, galls, and parasitoids. *Ecological Entomology*, **24**, 411-415.

Ferris C, King RA, Väinölä R & Hewitt GM (1998) Chloroplast DNA recognizes three refugial sources of European oaks and suggests independent eastern and western immigrations to Finland. *Heredity*, **80**, 584-593.

Ferris C, Oliver RP, Davy AJ & Hewitt GM (1993) Native oak chloroplasts reveal an ancient divide across Europe. *Molecular Ecology*, **2**, 337-344.

Ferris C, Oliver RP, Davy AJ & Hewitt GM (1995) Using chloroplast DNA to trace postglacial migration routes of oaks into Britain. *Molecular Ecology*, **4**,731-738.

Ferris, C., King, R. A., Väinölä, R. & Hewitt, G. M. (1998) Chloroplast DNA recognizes three refugial sources of European oaks and suggests independent eastern and western immigrations to Finland. *Heredity* **80**, 584-593.

Filatov DA (2002) ProSeq: a software for preparation and evolutionary analysis of DNA datasets. *Molecular Ecology Notes*, **2**, 621-624.

Fisher MC, Koenig GL, White TJ, San-Blas G, Negroni R, Alvarez IG, Wanke B & Taylor JW (2001). Biogeographic range expansion into South America by *Coccioides immitis* mirrors New World patterns of human migration. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 4558-4562.

Floyd R, Eyualem A, Papert A & Blaxter ML (2002) Molecular barcodes for soil nematode identification. *Molecular Ecology*, **11**, 839–850.

Folliot R (1964) Contribution a l' étude de la biologie des Cynipides gallicoles (Hyménoptères: Cynipoïdea). *Annales des Sciences Naturalles, Zoologie, Paris*, **12**, 407–564.

Folmer O, Black M, Hoeh W, LutzR & Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294-299.

Fonseca DM, Smith JL, Wilkerson RC & Fleischer RC (2006) pathways of expansion and multiple introductions illustrated by large genetic differentiation among worldwide populations of the southern house mosquito. *American Journal of Tropical Medicine and Hygiene*, **74**, 284-289.

Frati F, Simon C, Sullivan J & Swofford DL (1997) Gene evolution and phylogeny of the mitochondrial cytochrome oxidase gene in Collembola. *Journal of Molecular Evolution*, **44**, 145–158.

Frati F, Simon C, Sullivan J & Swofford DL (1997) Gene evolution and phylogeny of the mitochondrial cytochrome oxidase gene in Collembola. *Journal of Molecular Evolution*, **44**, 145–158.

Friedrich M & Tautz D (1995) Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution myriapods. *Nature*, **637**, 165-167.

Fry AJ & Zink RM (1998) Geographic analysis of nucleotide diversity and song sparrow (Aves: Emberizidae) population history. *Molecular Ecology*, **7**, 1303-1313.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **141**, 915-925

Gatesy J, Baker RH & Hayashi C (2004) Inconsistencies in arguments for the supertree approach: Supermatrices versus supertrees of Crocodylia. *Systematic Biology*, **53**, 342–355.

Gay L, Du Rau PD, Mondain-Monval JY & Crochet PA (2004) Phylogeography of a game species: the red-crested pochard (Netta rufina) and consequences for its management. *Molecular Ecology*, **13**, 1035-1045.

Geller JB, Walton ED, Grosholz ED. Ruiz GM (1997) Cryptic invasions of the crab *Carcinus* detected by molecular phylogeography. *Molecular Ecology*, **6**, 901-906.

Gelman A & Rubin DB (1992) Inference from iterative simulation using multiple sequences, *Statistical Science*, **7**, 457-511.

Geml J, Laursen GA, O'Neill K, Nusbaum HC & Taylor DL (2006) Beringian origins and cryptic speciation events in the fly agaric (*Amanita muscaria*). *Molecular Ecology*, **15**, 225-239.

Gibbard PL, Boreham S, Cohen KM & Moscariello A (2005) Global chronostratigraphical correlation table for the last 2.7 million years. *Boreas*, **34** (1), unpaginated, modified/updated 2007.

Goldman N (1993) Statistical tests of models of DNA substitution. *Journal of Molecular Evolution*, **36**, 182–198.

Govaerts R & Frodin DG (1998) *World checklist and bibliography of Fagales*. Royal Botanic Gardens, Kew. U.K.

Grapputo A, Boman S, Lindstrom L, Lyytinen A & Mappes J (2005) The voyage of an invasive species across continents: genetic diversity of North American and European Colorado potato beetle populations. *Molecular Ecology* **14**, 4207-4219.

Green DM (2005) Designatable units for status assessment of endangered species. *Conservation Biology*, **19**, 1813-1820.

Green DM, Sharbel TF, Kearsley J & Kaiser H (1996) Postglacial range fluctuation, genetic subdivision and speciation in the western North American spotted frog complex, *Rana pretiosa*. *Evolution*, **50**, 374-390.

Griffiths CS (1997) Correlation of functional domains and rates of nucleotide substitution in cytochrome *b*. *Molecular Phylogenetics and Evolution*, **7**, 352-365.

Gunduz I, Rambau RV, Tez C & Searle JB (2006) Mitochondrial DNA variation in the western house mouse (*Mus musculus domesticus*) close to its site of origin: studies in Turkey. *Biological Journal of the Linnean Society*, **84**, 473-485.

Guo SW & Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportions for multiple alleles. *Biometrics*, **48**, 361-372.

Gusfield D & Bansal V (2005) A fundamental decomposition theory for phylogenetic networks and incompatible characters. Pp. 217-232 In: Miyano S, Mesirov J, Kasif S, Istrail S, Pevzner P & Waterman M (eds.) Research in Computational Biology. Lecture Notes in Computer Science, vol. 3500, Springer-Verlag, Berlin.

Haag KL, Ayala FJ, Kamenetzky L, Gutierrez AM & Rosenzvit M (2004) Livestock trade history, geography and parasite strains: the mitochondrial genetic structure of *Echinococcus granulosus* in Argentina. *Journal of Parasitology*, **90**, 234-239.

Hachtel W (1980) Maternal inheritance of chloroplast DNA in some *Oenothera* species. *Journal of Heredity*, **71**, 191-194.

Hagemeijer WJM & Blair MJ (eds.) (1997) *The EBCC Atlas of European Breeding Birds*. Poyser, London, UK.

Hajibabaei M, De Waard JR, Ivanova NV, Ratnasingham S, Dooh RT, Kirk SL, Mackie PM & Hebert PDN (2005) Critical factors for assembling a high volume of DNA barcodes. *Philosophical transactions of the Royal Society of London, Series B*, **360**, 1959-1967.

Hancock JM, Tautz D & Dover GA (1988) Evolution of the secondary structures and compensatory mutations of the ribosomal RNAs of *Drosophila melanogaster*. *Molecular Biology and Evolution*, **5**, 393–414.

Hanson MR & Folkerts O (1992) Structure and function of the higher plant mitochondrial genome. *International Review of Cytology*, **141**, 129-172.

Hard JJ, Bradshaw WE & Holzapfel CM (1993) Genetic coordination of demography and phenology in the pitcherplant mosquito, *Wyeomyia smithii*. *Journal of Evolutionary Biology*, **6**, 707-723.

Harpending H (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology*, **66**, 591-600.

Harry M, Solignac M, Lachaie D (1998) Molecular evidence for parallel evolution of adaptive syndromes in fig-breeding *Lisocephala* (Drosophilidae). *Molecular Phylogenetics and Evolution*, **9**, 542-541.

Hartig T (1843) Zweiter nachtrag zur naturgeschichte der Gallwespen. *Zeitschrift für Entomologie* **4**, 395-422.

Hasegawa M, Kishino H & Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.*, **22**, 160-174.

Hayward A & Stone GN (2005) Oak gall wasp communities: evolution and ecology. *Basic and Applied Ecology*, **6**, 435-443.

Hayward A & Stone GN (2006) Comparative phylogeography across two trophic levels: the oak gall wasp *Andricus kollari* and its chalcid parasitoid *Megastigmus stigmatizans*. *Molecular Ecology*, **15**, 479-490.

Hebert PDN (1987) Genotypic characteristics of cyclic parthenogens and their obligately asexual derivatives. Pp. 175–218 in: *The evolution of sex and its consequences* (ed. Stearns, S.C.). Birkhauser Verlag, Basel, Switzerland.

Hebert PDN, Cywinska A, Ball SL & de Waard JR (2003) Biological identifications through DNA bar-codes. *Proceedings of the Royal Society of London, Series B*, **270**, 313-322.

Hebert PDN, Penton EH, Burns JM, Janzen DH & Hallwachs W (2004a) Ten species in one: DNA bar-coding reveals cryptic species in the neotropical skipper butterfly *Astrapes fugerator*. *Proceedings of the National Academy of Science USA*, **101**, 14812-14817.

Hebert PDN, Stoeckle MY, Zemlak TS & Francis CM (2004b) Identification of birds through DNA barcodes. *Public Library of Science: Biology*, **2**, 1657-1663.

Hernandez-Lopez A, Stone GN & Cook JM (2007) Molecular phylogeny and evolution of European oak galling wasps (Hymenoptera, Cynipidae) from mitochondrial and nuclear DNA sequences. Manuscript in preparation.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247-276.

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87-112.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B – Biological Sciences*, **359**, 183-195.

Hibbett DS, Gilbert LB & Donoghue MJ (2000) Evolutionary instability of ectomycorrhizal symbioses in basidiomycetes. *Nature*, **407**, 506–508.

Hoelzer GA, Wallman J & Melnick DJ (1998) The effects of social structure, geographical structure, and population size on the evolution of mitochondrial DNA: II. molecular clocks and the lineage sorting period. *Journal of Molecular Evolution*, **47**, 21–31.

Huber KT, Langton M, Penny D, Moulton V & Hendy M (2002) Spectronet: a package for computing spectra and median joining networks. *Applied Bioinformatics*, **1**, 2041-2059.

Huelsenbeck JP & Crandall KA (1997) Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Eology and Systematics*, **28**, 437–466.

Huelsenbeck JP & Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. *Bioinformatics*, **17**, 754-755.

Hundertmark KJ, Shields GF, Udina IG, Bowyer RT, Danilkin AA & Schwartz CC (2002) Mitochondrial phylogeography of moose (*Alces alces*): Late Pleistocene divergence and population expansion. *Molecular Phylogenetics and Evolution*, **22**, 375-387.

Huntley B & Birks HJ (1983) *An Atlas of Past and Present Pollen Maps for Europe, 0-13000 years ago*. Cambridge University Press, Cambridge.

Hurst GDD & Jiggins FM (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society of London, Series B*, **272**, 1525-1534.

Huson DH & Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254-267.

Huson DH & Klöpper TH (2005) Computing recombination networks from binary characters. *Bioinformatics*, **21**(suppl. 2), ii159-ii165.

Ibrahim KM, Nichols RA & Hewitt GM (1996) Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity*, **77**, 282-291

Irestedt M, Fjeldså J, Nylander JAA & Ericson P (2004) Phylogenetic relationships of typical Antbirds (Thamnophilidae) and test of incongruence based on Bayes factors. *BMC Evolutionary Biology*, **4**, 23.

Jaarola M & Searle JB (2002) Phylogeography of field voles (*Microtus agrestis*) in Eurasia inferred from mitochondrial DNA sequences. *Molecular Ecology*, **11**, 2613-2621.

Jalas J & Suominen J (eds.) (1987) *Atlas Florae Europaeae. Distribution of vascular plants in Europe* II. Angiosperms (part) 3: Salicaceae to Balanophoraceae; 4: Polygonaceae; 5: Chenopodiaceae to Basellaceae. Cambridge University Press, Cambridge.

Jeandroz S, Roy, A & Bousquet J (1997) Phylogeny and phylogeography of the circumpolar genus *Fraxinus* (Oleaceae) based on internal transcribed spacer sequences of nuclear ribosomal DNA. *Molecular Phylogenetics and Evolution*, **7**, 241-251.

Jermiin LS & Crozier RH (1994) The cytochrome b region in the mitochondrial DNA of the ant *Tetraponera rufoniger* sequence divergence in Hymenoptera may be associated with nucleotide content. *Journal of Molecular Evolution*, **38**, 282-294.

Johnson NK, Cicero C (2002) The role of ecological diversification in sibling speciation of *Empidonax* flycatchers (Tyrannidae): Multigene evidence from mtDNA. *Molecular Ecology*, **11**, 2065-2081.

Jukes TH & Cantor CR (1969) Evolution of protein molecules. In: HN Munro (ed.) *Mammalian protein metabolism*, pp. 21-132. Academic Press, New York, NY.

Juste J, Ibanez C, Munoz J, Trujillo D, Benda P, Karatas A, Ruedi M (2004) Mitochondrial phylogeography of the long-eared bats (*Plecotus*) in the Mediterranean Palaearctic and Atlantic islands. *Molecular Phylogenetics and Evolution*, **31**, 1114-1126.

Kass RE & Raftery AE (1995) Bayes factors. *Journal of the American Statistical Association*, **90**, 773–795.

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111-120.

King RA & Ferris C (1998) Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn. *Molecular Ecology*, **7**, 1151-1161.

Kinsey AC (1930) The gall wasp genus Cynips. A study in the origin of species. *Indiana University Studies*, **84-86**, 1-577.

Kinsey AC (1936) The origin of higher categories in Cynips. *Indiana University Publications. Science Series 4. Entomological Series*, **10**, 1-334.

Kinsey AC (1937) New Mexican gall wasps (Hymenoptera, Cynipidae). *Revista de Entomología*, **7**, 39-79.

Koch MA, Kiefer C, Ehrich D, Vogel J, Brochmann C & Mummenhoff K (2006) Three times out of Asia Minor: the phylogeography of *Arabis alpina* L. (Brassicaceae). *Molecular Ecology*, **15**, 825-839.

Konarov VL (ed.) (1936) *Flora of the USSR*. Izdatel'stva Akademii Nauk SSSR. Moscow- Leningrad, USSR.

Konnert M & Bergmann F (1995) The geographical distribution of genetic variation of silver fir (*Abies alba,* Pinaceae) in relation to its migration history. *Plant systematics and Evolution*, **196**, 19-30.

Kotlik P, Bogutskaya NG & Ekmekci FG (2004) Circum Black Sea phylogeography of *Barbus* freshwater fishes: divergence in the Pontic glacial refugium. *Molecular Ecology*, **13**, 87-95.

Kvacek Z & Walther H (1989) Palaeobotanical studies in Fagaceae of the European Tertiary. *Plant Systematics and Evolution*, **162**, 213-229.

Kvist L, Martens J, Ahola A & Orell M (2001) Phylogeography of a Palaearctic sedentary passerine, the willow tit (*Parus montanus*). *Journal of Evolutionary Biology*, **14**, 930-941.

Kwast E (2001) Range expansion of *Andricus aries* in Europe. *Cecidology*, **16**, 62–68.

Lagerkrantz U & Ryman N (1990) Genetic structure of Norway Spruce (*Picea abies*): concordance of morphological and allozymic variation. *Evolution*, **44**, 38-53.

Lambert DM, Baker A, Huynen L, Haddrath O, Hebert PDN & Millar CD (2005) Is a large-scale DNA-based inventory of ancient life possible? *Journal of Heredity*, **96**, 279-284.

Lawton JH, Bignell DE, Bolton B, Bloemers GF, Eggleton P, Hammond PM, Hodda M, Holt RD, Larsen TB, Mawdsley NA, Stork NE, Srivastava DS & Watt AD (1998) Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forests. *Nature*, **391**, 72–75.

Lee CT & Hastings A (2006) Non-equilibrium genetic structure is insensitive to the shape of the dispersal distribution. *Evolutionary Ecology Research*, **8**, 279-293.

Lehrmann T, Marcet PL, Graham DH, Dahl ER & Dubey JP (2006) Globalization and the population structure of *Toxoplasma gondii*. *Proceedings of the National Academy of Sciences of the USA*, **103**, 11423-11428.

Lenk P, Fritz U, Joger U & Wink M (1999) Mitochondrial phylogeography of the European pond turtle, *Emys orbicularis* (Linnaeus 1758). *Molecular Ecology*, **8**, 1911-1922.

Lessa EP, Cook JA & Patton JL (2003) Genetic footprints of demographic expansion in North America, but not Amazonia, during the Late Quaternary. *Proceedings of the National Academy of Sciences USA*, **100**, 10331-10334.

Liljeblad J & Ronquist F (1998) A phylogenetic analysis of higher-level gall wasp relationships (Hymenoptera: Cynipidae). *Systematic Entomology*, **23**, 229-252.

Liò P & Goldman N (1998) Models of Molecular Evolution and Phylogeny. *Genome Research*, **8**, 1233-1244.

Liukkonen-Antilla T, Uimaniemi L, Orell M & Lumme J (2002) Mitochondrial DNA variation and the phylogeography of the grey partridge (*Perdix perdix*) in Europe: from Pleistocene history to present day populations. *Journal of Evolutionary Biology*, **15**, 971-982.

Lopez P, Forterre P & Philippe H (1999) The root of the tree of life in the light of the covarion model. *Journal of Molecular Evolution*, **49**, 496–508.

Ludt CJ, Schroeder W, Rottmann O & Kuehn R (2004) Mitochondrial DNA phylogeography of red deer (*Cervus elaphus*). *Molecular Phylogeneics and Evolution*, **31**, 1064-1083.

Luikart G, Gielly L, Excoffier L, Vigne JD, Bouvet J & Taberlet P (2001) Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proceedings of the National Academy of Sciences USA*, **98**, 5927-5932.

Lumaret R, Tryphon-Dionnet M, Michaud H, Sanuy A, Ipotesi E, Born C & Mir C (2005) Phylogeographical variation of chloroplast DNA in Cork oak (*Quercus suber*). *Annals of Botany*, **96**, 853-861.

Magri D, Vendramin GG, Comps B, Dupanloup I, Geburek T, Gomory D, Latalowa M, Litt T, Paule L, Roure JM, Tantau I, van der Knaap WO, Petit RJ & de Beaulieu JL (2006) A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *New Phytologist*, **171**, 199-221.

Maisuradze NL (1962) Studies on the oak gall wasps in Great and Small Caucasus of Azarbaijan. *Scientific Notes of the Azarbaijan State University* **2**, 49-59 [In Russian].

Manchester SR (1981) Fossil plants of the Eocene Clarno Nut Beds. *Oregon Geologist*, **43**, 75-81.

Manos PS & Stanford AM (2001) The historical biogeography of Fagaceae: tracking the tertiary history of temperate and subtropical forests of the Northern Hemisphere. *International Journal of Plant Sciences*, **162**, 577-593, Supplement 6.

Manos PS, Doyle JJ & Nixon KC (1999) Phylogeny, Biogeography, and Processes of Molecular Differentiation in *Quercus* Subgenus *Quercus* (Fagaceae). *Molecular Phylogenetics and Evolution*, **12**, 333-349.

Matisoo-Smith E, Roberts RM, Irwin GJ, Allen JS, Penny D & Lambert DM (1998) Patterns of prehistoric human mobility in Polynesia indicated by mtDNA from the Pacific rat. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 15145-15150.

Matyas G & Sperisen C (2001) chloroplast DNA polymorphisms provide evidence for postglacial recolonisation of oaks (*Quercus* spp.) across the Swiss Alps. *Theoretical and Applied Genetics*, **102**, 12-20.

McDonald JH & Krietman M (1991) Adaptive protein evolution at the ADH locus in *Drosophila*. *Nature*, **351**, 651-654.

Meadows JRS, Li K, Kantanen J, Tapio M, Sipos W, Pardeshi V, Gupta V, Calvo JH, Whan V, Norris B & Kijas JW (2005) Mitochondrial sequence reveals high levels of gene flow between breeds of domestic sheep from Asia and Europe. *Journal of Heredity*, **96**, 494-501.

Médail F & Quézel P (1999) Biodiversity hotspots in the Mediterranean basin: Setting global conservation priorities. *Conservation Biology*, **13**, 1510-1513.

Melika G & Abrahamson WG (1997a) Descriptions of four new species of Cynipid gall wasps of the genus *Neuroterus* Hartig (Hym.: Cynipidae) with redescriptions of some known species from the eastern United States. *Proceedings of the Entomological Society of Washington*, **99**, 560-573.

Melika G & Abrahamson WG (1997b) Synonymy of two genera (*Eumayria* and *Trisoleniella*) of Cynipid gall wasps and description of a new genus, *Eumayriella* (Hymenoptera: Cynipidae). *Proceedings of the Entomological Society of Washington*, **99**, 666-675.

Melika G & Abrahamson WG (2000) Review of the cynipid gall wasps of the genus *Loxaulus* Mayr (Hymenoptera: Cynipidae) with descriptions of new species. *Proceedings of the Entomological Society of Washington*, **102**, 198-211.

Melika G & Bechtold M (2001) Taxonomic notes and type designations of gall inducing cynipid wasps described by G. Mayr (Insecta: Hymenoptera: Cynipidae). *Annalen des Naturhistorischen Museums in Wien*, **103B**, 327-339.

Melika G & Stone GN (2001) *Andricus askewi*: a new species of Cynipid gall wasp from Turkey (Hymenoptera: Cynipidae). *Folia Entomologica Hungarica*, **62**, 127-131.

Melika G, Csóka G & Pujade-Villar J (2000) Checklist of oak gall wasps of Hungary, with some taxonomic notes (Hymenoptera: Cynipidae, Cynipinae, Cynipini). *Annales Historico-Naturalis Musei Nationalis Hungarici*, **92**, 265-296.

Melika G, Ros-Farré P, Pénzes Zs, Ács Z & Pujade-Villar J (2005) *Ufo abei* Melika et Pujade-Villar (Hymenoptera: Cynipidae: Synergini) new genus and new species from Japan. *Acta Zoologica Academiae Scientiarum Hungaricae*, **51**, 313-327.

Melika G, Sadeghi SE, Tavakoli M, Pénzes Z, Challis RJ, Assareh MA, Atkinson R, Bechtold M, Mikó I, Zargaran MR, Aligolizade D, Barimani H & Stone GN (2007) New Species of Oak Gallwaps from Iran (Hymenoptera: Cynipidae: Cynipini). *Zootaxa*, submitted manuscript.

Melika G, Stone GN & Csóka G (1999) Description of an oak gall-wasp, *Neuroterus ambrusi* sp. n. (Hymenoptera, Cynipidae) from Hungary. *Acta Zoologica Academiae Scientiarum Hungaricae*, **45**, 335-343.

Merilä J, Björklund M & Baker AJ (1997) Historical demography and present day population structure of the Greenfinch, *Carduelis chloris* - an analysis of mtDNA control-region sequences. *Evolution*, **51**, 946-956.

Metropolis N, Rosenbluth A, Rosenbluth M, Teller A & Teller E (1953) Equations of state calculations by fast computing machines. *Journal of Chemical Physics.*, **21**, 1087-1092.

Metzlaff M, Borner T & Hagemann (1981) Variations of chloroplast DNAs in the genus *Pelargonium* and their biparental inheritance. *Theoretical and Applied Genetics*, **60**, 37-41.

Michaux JR, Libois R & Filippucci M-G (2005) So close and yet so different: comparative phylogeography of two small mammal species, the Yellow-necked fieldmouse (*Apodemus flavicollis*) and the Woodmouse (*Apodemus sylvaticus*) in the Western Palaearctic region. *Heredity*, **94**, 52-63.

Michaux JR, Magnanou E, Paradis E, Nieberding C, Libois R (2003) Mitochondrial phylogeography of the woodmouse (*Apodemus sylvaticus*) in the Western Palaearctic region. *Molecular Ecology*, **12**, 685-697.

Mitchell-Jones AJ, Amori G, Bogdanowicz W, Kryštufek B, Reijnders PJH, Spitzenberger F, Stubbe M, Thissen JBM, Vohralík V & Zima J (1999) *The Atlas of European Mammals*. Poyser, London, UK.

Morgan JAT, Dejong RJ, Adeoye GO, Ansa EDO, Barbosa CS, Bremond P, Cesari IM, Charbonnel N, Correa LR, Coulibaly G, D'Andrea PS, De Souza CP, Doenhoff MJ, File S, Idris MA, Incani RN, Jarne P, Karanja DMS, Kazibwe F, Kpikpi J, Lwambo NJS, Mabaye A, Magalhaes LA, Makundi A, Mone H, Mouahid G, Muchemi GM, Mungai BN, Sene M, Southgate V, Tchuente LAT, Theron A, Yousif F, Magalhaes EMZ, Mkoji GM, Loker ES & Loker ES (2005) Origin and diversification of the human parasite *Schistosoma mansoni*. *Molecular Ecology*, **14**, 3889-3902.

Morii Y, Ishiguro N, Watanabe T, Nakano M, Hongo H, Matsui A & Nishimoto T (2002) Ancient DNA reveals genetic lineage of *Sus scrofa* among archaeological sites in Japan. *Anthropological Science*, **110**, 313-328.

Moritz C & Cicero C (2004) DNA bar-coding: Promise and pitfalls. *Public Library of Science: Biology*, **2**, 1529-1531.

Mousson L, Dauga C, Garrigues T, Schaffner F, Vazeile M & Failloux AB (2005) Phylogeography of *Aedes* (*Stegomyia*) *aegypti* (L.) and *Aedes* (Stegomyia) *albopictus* (Skuse) (Diptera: Culicidae) based on mitochondrial DNA variations. *Genetical Research*, **86**, 1-11.

Muller CJ (1981) Fossil records of extant angiosperms. *Botanical Review*, **47**, 1-142.

Nardi F, Carapelli A, Dallai R, Roderick GK, Frati F (2005) Population structure and colonization history of the olive fly, *Bactrocera oleae* (Diptera, Tephritidae). *Molecular Ecology*, **14**, 2729-2738.

Navia D, de Moraes GJ, Roderick G & Navajas M (2005) The invasive coconut mite *Aceria guerreronis* (Acari: Eriophyidae): origin and invasion sources inferred from mitochondrial (16S) and nuclear (ITS) sequences. *Bulletin of Entomological Research*, **95**, 505-516.

Nee S (2001) Inferring speciation rates from phylogenies. *Evolution*, **55**, 661-668.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nesbø CL, Fosshein T, Vøllestad LA & Jakobsen KS (1999) Genetic divergence and phylogeographic relationships among European perch (*Perca fluviatilis*) populations reflect glacial refugia and postglacial colonization. *Molecular Ecology*, **8**, 1387-1404.

Niebuhr C (1776-1780) *Reize naar Arabië en andere omliggende Landen*. J.S. Baalde, Amsterdam.

Nieves Aldrey J-L (1992) Revision de las especies Europeas del genero *Callirhytis* Forster (Hymenoptera, Cynipidae). *Graellsia*, **48**, 171-183.

Nieves Aldrey JL (2001) *Fauna Iberica*. Hymenoptera, Cynipidae. Museo Nacional de Ciencias Naturales Consejo Superior de Investigaciones Cientificas, Madrid. Vol. **16**, 1-636.

Nieves-Aldrey JL (1983a) Sobre las especies de Sycophila Walker, asociadas con agallas de cinipidos en la Peninsula Iberica, con descripcion de una nueva especie (Hym., Eurytomidae). *Eos*, **59**, 179–191.

Nieves-Aldrey JL (1983b) Sobre las especies del genero Mesopolobus (Hym., Pteromalidae) asociadas con agallas de cynipidos en *Quercus sp.* en Salamanca. *Boletín de la Asociacíon Española de Entomología*, **7**, 9–18.

Nieves-Aldrey JL (1983c) Contribucion al conocimiento de los eulófidos (Hym., Chalcidoidea, Eulophidae) parásitos en las agallas de cinipidos producidas sobre especies de *Quercus*. *Boletín de la Asociacíon Española de Entomología*, **6**, 43–54.

Nieves-Aldrey JL (1984a) Primeros datos sobre los representantes de la familia Ormyridae en Espana, con descripción de una nueva especie. *Graellsia*, **40**, 119–127.

Nieves-Aldrey JL (1984b) Observaciones sobre los Torimidos (Hym., Chalcidoidea, Torymidae) asociados con agallas de Cinipidos (Hym., Cynipidae) sobre *Quercus spp.* en la zona centro-occidental de Espana. *Boletín de la Asociacíon Española de Entomología*, **8**, 121–134.

Nieves-Aldrey JL (2001) Hymenoptera; Cynipidae. Fauna Ibérica. Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Cientificas, Madrid, Spain.

Nieves-Aldrey JL, & Pujade-Villar J (1985) Sobre las especies ibéricas de la Sección 1 (Mayr, 1872) del género *Synergus* Htg. *Eos*, **61**, 219–237.

Nieves-Aldrey JL, & Pujade-Villar J (1986) Sobre las especies ibéricas de la Sección II (Mayr, 1872) del género *Synergus* Htg. con descripción de una especie nueva (Hym., Cynipidae, Cynipinae). *Eos*, **62**, 137-165.

Nieves-Aldrey J-L, Pujade-Villar J, Schönrogge K & Melika G (2007) Cynipid inquilines of oak gall wasps. In: Stone GN, Csóka G, Melika G (eds.) *The oak gallwasps of the Western Palaearctic: ecology, evolution and systematics*. The Ray Society, London.

Nieves-Aldrey JL, Vårdal H & Ronquist F (2005) Comparative morphology of terminal-instar larvae of Cynipoidea: phylogenetic implications. *Zoologica Scripta* **34**, 15-36.

Nixon KC (1993) Infrageneric classification of *Quercus* (Fagaceae) and typification of sectional names. *Annales des Sciences Forestières*, **50**, Supplement 25-34.

Nordlander G (1984) What do we know about parasitic cynipoids (Hymenoptera). *Entomologisk Tidskrift*, **105**, 36-40.

Nylander JAA (2004) MrModeltest v2. Evolutionary Biology Centre, Uppsala University.

Nylander JAA, Buffington ML, Liu Z, Nieves-Aldrey JL, Liljeblad J & Ronquist F (2004a) Molecular Phylogeny and Evolution of Gall Wasps. In: Nylander JAA *Bayesian Phylogenetics and the Evolution of Gall Wasps*, PhD Thesis, University of Uppsala, Uppsala, Sweden.

Nylander JAA, Ronquist F, Huelsenbeck JP & Nieves Aldrey J-L (2004b). Bayesian phylogenetic analysis of combined data. *Systematic Biology*, **53**, 47-67.

Orth A, Auffrey JC & Bonhomme F (2002) Two deeply divergent mitochondrial clades in the wild mouse *Mus macedonicus* reveal multiple glacial refuges south of Caucasus. *Heredity*, **89**, 353-357.

Pace NR (1997) A molecular view of microbial diversity and the biosphere. *Science*, **276**, 734–740.

Pagel M & Meade A (2004) A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Systematic Biology*, **53**, 571-581.

Pagel M (1999) The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Systematic Biology*, **48**, 612-622.

Pagel M, Meade A & Barker D (2004) Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology*, **53**, 673-684.

Palmer JD & Herbon LA (1988) Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *Journal of Molecular Evolution*. **28**, 87-92.

Papakostas S, Triantafyllidis A, Kappas I & Abatzopoulos TJ (2005) The utility of the 16S gene in investigating cryptic speciation within the *Brachionus plicatilis* species complex. *Marine Biology*, **147**, 1129-1139.

Pavlova A, Rohwer S, Drovetski SV & Zink RM (2006) Different post-pleistocene histories of Eurasian parids. *Journal of Heredity*, **97**, 389-402.

Peakall R & Beattie AJ (1991) The genetic consequences of worker ant pollination in a self-compatible clonal orchid. *Evolution*, **45**,1837-1848.

Petit RJ, Csaikl UM, Bordacs S, Burg K, Coart E, Cottrell J, van Dam B, Deans JD, Dumolin-Lapegue S, Fineschi S, Finkeldey R, Gillies A, Glaz I, Goicoechea PG, Jensen JS, Konig AO, Lowe AJ, Madsen SF, Matyas G, Munro RC, Olalde M, Pemonge MH, Popescu F, Slade D, Tabbener H, Taurchini D, de Vries SGM, Ziegenhagen B & Kremer A (2002) Chloroplast DNA variation in European white oaks - Phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management*, **156**, 5-26.

Pfenninger M, Posada D (2002) Phylogeographic history of the land snail *Candidula unifasciata* (Hellicellinae, Stylommatophora): fragmentation, corridor migration, and secondary contact. *Evolution*, **56**, 1776-1788.

Philippe H & Forterre P (1999) The rooting of the universal tree of life is not reliable. *Journal of Molecular Evolution*, **49**, 509–523.

Philippe H, Lartillot N & Brinkmann H (2005) Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Molecular Biology and Evolution*, **22**, 1246-1253.

Philippe H, Snell EA, Bapteste E, Lopez P, Holland PWH & Casane D (2004) Phylogenomics of Eukaryotes: Impact of Missing Data on Large Alignments. *Molecular Biology and Evolution*, **21**, 1740-1752.

Plantard O, Hochberg ME (1998) Factors affecting parasitism in the oak galler *Neuroterus quercusbaccarum* (Hymenoptera: Cynipidae). *Oikos*, **81**, 289–98.

Plantard O, Rasplus JY, Hochberg ME (1996) Resource partitioning in the parasitoid assemblage of the oak galler *Neuroterus quercusbaccarum* L. (Hymenoptera: Cynipidae). *Acta Oecologica*, **17**, 1–15.

Pons AO & Quézel P (1985) The history of the flora and vegetation and past and present human disturbance in the mediterranean region. Pp 25-43 in: *Plant Conservation in the Mediterranean Area* (ed. C. Gomez-Campo). W. Junk, Dordrecht, The Netherlands.

Posada D & Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. *Systematic Biology*, **53**, 793-808.

Posada D & Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics*, **14**, 817-818.

Posada D & Crandall KA (2001) Selecting the best-fit model of nucleotide substitution. *Systematic Biology*, **50**, 580-601.

Posada D, Crandall KA, Templeton AR (2000) GEODIS: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology*, **9**, 487-488.

Powell JR (1983) Interspecific cytoplasmic gene flow in the absence of nuclear gene flow: evidence from *Drosophila*, *Proceedings of the National Academy of Sciences USA*, **80**, 492-495.

Powers T (2004) Nematode molecular diagnostics: from bands to barcodes. *Annual Review of Phytopathology*, **42**, 367-383.

Pritchard JK, Stephens M & Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.

Pujade-Villar J & Nieves-Aldrey JL (1990) Revisión de las especies europeas del género *Saphonecrus* Dalla Torre and Kieffer, 1910 (Hymenoptera: Cynipidae: Cynipinae). *Bulletín de la Institució Catalana d'Historia Natural*, **58**, 45–55.

Pujade-Villar J & Nieves-Aldrey JL (1993) Revisión de las especies europeas del género *Ceroptres* Htg. 1840 (Hymenoptera: Cynipidae). *Boletin de la Asociación Española de Entomologia*, **17**, 49–63.

Pujade-Villar J (1989) Sobre la variabilitat morfologia d'algunes espècies dels Ormyridae Foerster, una nueva familia d'himenopters per a Catalunya (Hymenoptera, Chalcidoidea). *Sessió Conjunta d'Entomologia de l'institució Catalanya d'Hostória Natural (ICHN) I la Societat Catalana de Lepideroptologia (SCL)*, **5**, 77-90.

Pujade-Villar J (1992a) *Andricus kollari* (Htg.) (Insecta: Hymenoptera: Cynipidae) 2ª part: Consideracions sobre seu cicle biologic. *La Sitja del Llop*, **3**, 12.

Pujade-Villar J (1992b) Sobre la presencia de *Hobbya* Delucchi, 1957 en el nordeste Ibérico (Hym., Pteromalidae). *Eos*, **68**, 201-202.

Pujade-Villar J (1992c) Sobre *Synergus ilicinus* (Barbotin, 1972) n. comb. y otras especies de *Synergus* Htg. en el Nordeste de la Peninsula Ibérica (Hym., Cynipidae, Synergini). *Boletín de la Asociación española de Entomología*, **16**, 129-148.

Pujade-Villar J, Bellido D, Segú G & Melika G (2001) Current state of knowledge of heterogeny in Cynipidae (Hymenoptera, Cynipoidea). *Sessió Conjunta d' Entomologia de l'Institució Catalana d' História Natural i la Societat Catalana de Lepideroptologia* **11**, 97-107.

Pujade-Villar J, Folliot R & Bellido D (2003) The life cycle of *Andricus hispanicus* (Hartig, 1856) n. stat., a sibling species of *A. kollari* (Hartig, 1843) (Hymenoptera: Cynipidae). *Butlletí de l'Institució Catalana d'História Natural*, **71**, 83–95.

Pujade-Villar J, Kovalev O & Liljeblad J (2003) Status of *Pseudoneuroterus* Kinsey, 1923 and remarks on closely related genera (Hymenoptera, Cynipidae). *Nouveau Revue Entomologique* (*nouvelle serie*), **20**, 353-360.

Pujade-Villar J, Kwast E, Thuróczy Cs & Bellido D (2002) Gall wasps and their associated fauna collected in Greece, with some taxonomic changes and description of a new species (Hymenoptera: Cynipidae, Chalcidoidea). *Acta Zoologica Cracovensia, Krakow*, **45**, 351-364.

Pujade-Villar J, Melika G, Ros-Farré P, Ács Z & Csóka Gy (2003) Cynipid inquiline wasps of Hungary, with taxonomic notes on the Western Palaearctic fauna (Hymenoptera: Cynipidae, Cynipinae, Synergini). *Folia Entomologia Hungarica*, **64**, 121-170.

Pybus OG (2006) Model selection and the molecular clock. *Public Library of Science Biol.* **4**:686-688.

Quicke D (2004) The world of DNA bar-coding and morphology – collision or synergism and what of the future? *The Systematist*, **23**, 8-12.

Quinlan J (1974) On the occurrence of *Andricus lignicola* (Hartig) (Hym., Cynipidae) in Britain. *The Entomologist's Gazette*, **25**, 293-296.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092–2100.

Rannala B, Huelsenbeck JP, Yang Z & Nielsen R (1998) Taxon sampling and the accuracy of large phylogenies. *Systematic Biology*, **47**, 702–710.

Richardson BJ, Baverstock PR & Adams M (1986) *Allozyme Electrophoresis; a Handbook for Animal Systematics and Population Studies*. Academic Press. New York.

Ritchie AJ & Shorthouse JD (1987) A review of the species of *Synergus* from Guatemala, with notes on *Cynips guatemalensis* Cameron (Hymenoptera: Cynipidae). *Proceedings of the Entomological Society of Washington*, **89**, 230-241.

Roderick GK (1996) Geographic structure of insect populations: gene flow, phylogeography, and their uses. *Annual Review of Entomology*, **41**, 325-352.

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552-569.

Rokas A, Atkinson RJ, Brown G, West SA & Stone GN (2001) Understanding patterns of genetic diversity in the oak gall wasp *Biorhiza pallida*: demographic history or a *Wolbachia* selective sweep? *Heredity*, **87**, 294–305.

Rokas A, Atkinson RJ, Nieves-Aldrey JL, West SA, & Stone GN (2002a) The incidence and diversity of *Wolbachia* in gallwasps (Hymenoptera; Cynipidae) on oak. *Molecular Ecology*, **11**, 1815-1829.

Rokas A, Atkinson RJ, Webster L, Csóka G & Stone GN (2003a) Out of Anatolia: longitudinal gradients in genetic diversity support an eastern origin for a circum-Mediterranean oak gallwasp *Andricus quercustozae*. *Molecular Ecology*, **12**, 2153-2174.

Rokas A, Melika G, Abe Y, Nieves-Aldrey J-L, Cook JM & Stone GN (2003b) Lifecycle closure, lineage sorting, and hybridisation revealed in a phylogenetic analysis of European oak gallwasps (Hymenoptera: Cynipidae: Cynipini) using mitochondrial sequence data. *Molecular Phylogenetics and Evolution*, **26**, 36-45.

Rokas A, Nylander JAA, Ronquist F & Stone GN (2002b) A maximum-likelihood analysis of eight phylogenetic markers in gallwasps (Hymenoptera: Cynipidae): implications for insect phylogenetic studies. *Molecular Phylogenetics and Evolution*, **22**, 206-219.

Rokas A, Williams BL, King N & Carroll SB (2003c) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*, **425**, 798-803.

Ronquist F & Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572-1574.

Ronquist F & Liljeblad J (2001) Evolution of the gall wasp-host-plant association. *Evolution*, **55**, 2503–2522.

Ronquist F (1994) Evolution of parasitism among closely related species: phylogenetic relationships and the origin of inquilinism in gall wasps (Hymenoptera, Cynipidae). *Evolution*, **48**, 241-266.

Ronquist F (1995). Phylogeny and early evolution of the Cynipoidea (Hymenoptera). *Systematic Entomology*, **20**, 309-335.

Ronquist F (1999) Phylogeny of the Hymenoptera (Insecta): The state of the art. *Zoologica Scripta*, **28**, 3-11.

Rozas J, Sanchez-Del Barrio JC, Messeguer X, Rozas R (2003) DNASP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496-2497

Sadeghi SE, Melika G, Pujade-Villar J, Pénzes Z, Ács Z, Bechtold M, Assareh MH, Tavakoli M, Yarmand H, Askary H, Stone GN, Azizkhani E, Zargaran MR, Aligolizade D, Barimani H & Dordaei AA (2006) Oak cynipid gall inquilines of Iran (Hym.: Cynipidae: Synergini), with descriptions of new species. *Journal of the Entomological Society of Iran,* **25**, 15-50.

Sage RD & Wolff JO (1986) Pleistocene glaciations, fluctuating ranges, and low genetic variability in a large mammal (*Ovis dalli*). *Evolution*, **40**, 1092-1095.

Saitou N & Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406-425.

Sanderson AR (1988) Cytological investigations of parthenogenesis in gall wasps (Cynipidae, Hymenoptera). *Genetica*, **77**, 189–216.

Sanderson MJ & Kim J (2000) Parametric phylogenetics? *Systematic Biology*, **49**, 817–829.

Sanmartin I, Enghof H & Ronquist F (2001) Patterns of animal dispersal, vicariance and diversification in the Holarctic. *Biological Journal of the Linnean Society*, **73**, 345-390.

Sanver D & Hawkins BA (2000) Galls as habitats: the inquiline communities of insect galls. *Basic and Applied Ecology*, **1**, 3-11.

Schaal BA, Hayworth BA, Olsen KM, Rauscher JT & Smith WA (1998) Phylogeographic studies in plants: problems and prospects. *Molecular Ecology*, **7**, 465-474.

Schmidt HA, Strimmer K, Vingron M & von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*, **18**, 502-504.

Schmitt T, Röber S, Seitz A (2005) Is the last glaciation the only relevant event for the present genetic population structure of the meadow brown butterfly *Maniola jurtina* (Lepidoptera: Nymphalidae)? *Biological Journal of the Linnean Society,* **85**, 419-431.

Schneider S & Excoffier L (1999) Estimation of demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: Application to human mitochondrial DNA. *Genetics*, **152**, 1079-1089.

Schneider S, Roessli D & Excoffier L (2000) Arlequin ver. 2.0: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Schönrogge K & Askew RR (2007) Oak cynipid communities. In: Stone GN, Csóka G, Melika G (eds.) *The oak gallwasps of the Western Palaearctic: ecology, evolution and systematics*. The Ray Society, London.

Schonrogge K, Crawley MJ (2000) Quantitative webs as a means of assessing the impact of alien insects. *Journal of Animal Ecology*, **69**, 841-868.

Schönrogge K, Stone GN & Crawley MJ (1994) Spatial and temporal variation in guild structure: parasitoids and inquilines of *Andricus quercuscalicis* (Hymenoptera: Cynipidae) in its native and alien ranges. *Oikos*, **72**, 51–60.

Schönrogge K, Stone GN & Crawley MJ (1996a) Alien herbivores and native parasitoids: rapid development of guild structure in an invading gall wasp, *Andricus quercuscalicis* (Hymenoptera: Cynipidae). *Ecological Entomology,* **21**, 71–80.

Schönrogge K, Stone GN & Crawley MJ (1996b) Abundance patterns and species richness of the parasitoids and inquilines of the alien gall former *Andricus quercuscalicis* Burgsdorf (Hymenoptera: Cynipidae). *Oikos*, **77**, 507–518.

Schönrogge K, Walker P & Crawley MJ (1998) Invaders on the move: parasitism in the galls of four alien gall wasps in Britain (Hymenoptera, Cynipidae). *Proceedings of the Royal Society of London, Series B*, **256**, 1643-1650.

Schönrogge K, Walker P & Crawley MJ (1999) Complex life cycles in *Andricus kollari* (Hymenoptera, Cynipidae) and their impact on associated parasitoid and inquiline species. *Oikos*, **84**, 293–301.

Schwartz G (1978) Estimating the dimension of a model. *Annals of Statistics*, **6**, 461-464.

Schwartz O (1964) *Quercus* L. Pp 72-76 In: Tutin, T.G., Heywood VH, Burges NA, Morre DM, Valentine DH, Walters SM & Webb DA (eds.) *Flora Europaea*, 1st edition, vol 1. Cambridge University Press, Cambridge.

Seibert TF (1993) A nectar secreting gall wasp and ant mutualism – selection and counter-selection shaping gall wasp phenology, fecundity and persistence. *Ecological entomology*, **18**, 247-253.

Shmida A (1984) Endemism in the flora of Israel. *Botanische Jahrbucher für Systematik, Pflanzengeschichte und Pflanzengeographie*, **104**, 537-567.

Sitch TA, Grewcock, DA & Gilbert FS (1988) Factors affecting components of fitness in a gallmaking wasp (*Cynips divisa* Hartig). *Oecologia*, **76**, 371-375.

Slatkin M & Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, **129**, 555-562.

Smith MA, Fisher BL & Hebert PDN (2005) DNA bar-coding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Philosophical transactions of the Royal Society of London, Series B*, **360**, 1825-1834.

Sokal RR & Rohlf FJ (1981) *Biometry*. WH Freeman, New York.

Sokal RR, Oden NL & Wilson C (1991) Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature*, **351**, 143-145.

Sommer RS & Benecke N (2005) The recolonization of Europe by brown bears *Ursus arctos* Linnaeus, 1758 after the Last Glacial Maximum. *Mammal Review*, **35**, 156-164.

Sota T, Ishikawa R, Ujie M, Kusumoto F & Vogler AP (2001) Extensive trans-species mitochondrial polymorphisms in the carabid beetles *Carabus* subgenus *Ohomopterus* caused by repeated introgressive hybridisation. *Molecular Ecology*, **10**, 761-778.

Starrett J & Hedin M (2007) Multilocus genealogies reveal multiple cryptic species and biogeographical complexity in the California turret spider *Antrodiaetus riversi* (Mygalomorphae, Antrodiaetidae). *Molecular Ecology*, **16**, 583-604.

Sternlicht M (1968) The Oak galls of Israel (*Quercus calliprinos* Webb, and *Quercus ithaburensis* Decne.) *Volcani Inst. of Agricultural Research* **No. 1198-E**, 17-44.

Stone G, Atkinson R, Rokas A, Csóka G & Nieves-Aldrey J-L (2001) Differential success in northwards range expansion between ecotypes of the marble gallwasp: *Andricus kollari*: a tale of two lifecycles. *Molecular Ecology*, **10**, 761-778.

Stone GN & Cook JM (1998) The structure of cynipid oak galls: patterns in the evolution of an extended phenotype. *Proceedings of the Royal Society of London, Series B*, **265**, 979-988.

Stone GN & Schönrogge K (2003) The adaptive significance of insect gall morphology. *Trends in Ecology and Evolution*, **18**, 512-522.

Stone GN & Sunnucks PJ (1992) The hedgehog gall *Andricus lucidus* (Hartig 1843) confirmed in Britain. *Cecidology*, **7**, 30–35.

Stone GN & Sunnucks P (1993) Genetic consequences of an invasion through a patchy environment - the Cynipid gallwasp *Andricus quercuscalicis* (Hymenoptera, Cynipidae). *Molecular Ecology*, **2**, 251-268.

Stone GN, Atkinson RJ & Csóka G (2007a) Host plant associations of oak gallwasps. In: Stone GN, Csóka G, Melika G (eds.) *The oak gallwasps of the Western Palaearctic: ecology, evolution and systematics*. The Ray Society, London.

Stone GN, Atkinson RJ, Melika G & Bordacs S (2007b) The biogeography and origins of western palaearctic oak gallwasp communities. In: Stone GN, Csóka G, Melika G (eds.) *The oak gallwasps of the Western Palaearctic: ecology, evolution and systematics*. The Ray Society, London.

Stone GN, Atkinson RJ, Rokas A, Brown G & Csóka G (2002a) The Population Genetic Consequences of Range Expansion. Pp. 46-62 In: Hails RS, Beringer JE & Godfray HCJ (eds.) *Genes in the environment*. Special Symposium Volume, The British Ecological Society. Blackwells Science.

Stone GN, Challis RJ, Atkinson RJ, Csóka G, Hayward A, Mutun S, Preuss S, Rokas A, Sadeghi E & Schönrogge K (2007c) The phylogeographic clade trade: Tracing the impact of human-mediated dispersal on the colonisation of Northern Europe by the oak gallwasp *Andricus kollari*. *Molecular Ecology*, in press.

Stone GN & Cook JM (1998) The structure of cynipid oak galls: patterns in the evolution of an extended phenotype. *Proceedings of the Royal Society of London, Series B*, **265**, 979-988.

Stone GN, Melika G, Pujade-Villar J, Nieves-Aldrey JL, Csóka G, Folliot R & Plantard O (2007d) Summaries of the biology of all Western Palaearctic oak cynipids. In: Stone GN, Csóka G, Melika G (eds.) *The oak gallwasps of the Western Palaearctic: ecology, evolution and systematics*. The Ray Society, London, in press.

Stone GN, Schönrogge K, Atkinson RJ, Bellido D & Pujade-Villar J (2002b) The population biology of oak gall wasps (Hymenoptera: Cynipidae). *Annual Review of Entomology*, **47**, 633-668.

Strimmer K & Moulton V (2000) Likelihood Analysis of Phylogenetic Networks Using Directed Graphical Models. *Molecular Biology and Evolution*, **17**, 875-881.

Strimmer K & von Haeseler A (1997) Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment. *Proceedings of the National Academy of Sciences USA*, **94**, 6815-6819.

Swofford DL (2001) PAUP*. *Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates, Sunderland, MA.

Swofford DL, Waddell PJ, Huelsenback JP, Foster PG, Lewis PO & Rogers JS (2001) Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Systematic Biology*, **50**, 525-539.

Taberlet P & Bouvet J (1994) Mitochondrial-DNA polymorphism, phylogeography, and conservation genetics of the brown bear *Ursus arctos* in Europe. *Proceedings of the Royal Society of London, Series B*, **255**, 195-200.

Taberlet P, Fumagalli L, Wust-Saucy A-G & Cosson J-F (1998) Comparative phylogeography and postglacial colonisation routes in Europe. *Molecular Ecology*, **7**, 453-464.

Takhtajan A (1986) *Floristic Regions of the World*. California University Press, Berkely, California, USA.

Tavaré S (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. In: RM Miura (ed.) *Lectures in mathematics in the life sciences*, Vol. 17, pp. 57-86. American Mathematical Society, Providence, RI.

Templeton AR (1998) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology*, **7**, 381-397.

Templeton AR (2004) Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology*, **13**, 789-809.

Templeton AR, Boerwinkle E, Sing CF (1987) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics*, **117**, 343-351.

Templeton AR, Routman E, Phillips C (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics*, **140**, 767-782.

Templeton AR & Sing CF (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics*, **134**, 659-669.

Thelwell NJ, Huisman RA, Harbach RE, Butlin RK (2000) Evidence for

mitochondrial introgression between *Anopheles bwambae* and *Anopheles gambiae*. *Insect Molecular Biology*, **9**, 203-210.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F & Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **24**, 4876-4882.

Thorne JL, Kishino H & Painter IS (1998) Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution*, **15**, 1647-1657.

Tierney L (1994) Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics*, **22**, 1701-1762.

Tomaru N, Mitsutsuji T, Takahashi M, Tsumura Y, Uchida K & Ohba K (1997) Genetic diversity in *Fagus crenata* (Japanese Beech): influence of the distributional shift during the late Quaternary. *Heredity*, **78**, 241-251.

Townsend CC & Guest ER (eds.) (1980) *Flora of Iraq*. Volume 4, part 1: Cornaceae to Rubiaceae. Ministry of Agriculture and Agrarian Research, Baghdad, Iraq.

Trelease W (1924) The American oaks. *Memoirs of the National Academy of Sciences*, **20**, 1-255.

Triantafyllidis A, Krieg F, Cottin C, Abatzopoulos TJ, Triantaphyllidis C & Guyomard R (2002) Genetic structure and phylogeography of European catfish (*Silurus glanis*) populations. *Molecular Ecology*, **11**, 1039-1055.

Tutin TG, Heywood VH, Burges NA, Moore DM, Valentine DH, Walters SM & Webb DA (1993) *Flora Europaea*. Cambridge University Press, London, U.K.

Tzedakis PC (1993) Long-term tree populations in northwest Greece through multiple quaternary climatic cycles. *Nature* **364**, 437-440.

Ursenbacher S, Carlsson M, Helfer V, Tegelstrom H & Fumagalli L (2006) Phylogeography and Pleistocene refugia of the adder (*Vipera berus*) as inferred from mitochondrial DNA sequence data. *Molecular Ecology*, **15**, 3425-3437.

van Noort S, Stone GN, Whitehead VB & Nieves-Aldrey JL (2006) Biology and redescription of *Rhoophilus loewi* (Cynipidae: Cynipoidea: Hymenoptera), with evolutionary implications on the inquilinism in gall wasps. *Biological Journal of the Linnean Society*, 90, 153-172.

Vandewoestijne S, Baguette M, Brakefield PM & Saccheri IJ (2004) Phylogeography of *Aglais urticae* (Lepidoptera) based on DNA sequences of the mitochondrial COI gene and control region. *Molecular Phylogenetics and Evolution*, **31**, 630-646.

Wagner DB, Furnier GR, Saghai-Maroof MA, Williams SM, Dancik BP & Allard RW (1987) Chloroplast DNA polymorphisms in lodgepole and jack pines and their hybrids. *Proceedings of the National Academy of Science USA*, **84**, 2097-2100.

Walker P (2001) *Andricus aries* does have a sexual generation on Turkey oak. *Cecidology*, **16**, 94–95.

Walker P (2002) Two new records for cynipid oak galls in Britain. *Cecidology*, **17**, 64.

Walker P, Leather SR & Crawley MJ (2002) Differential rates of invasion in three related alien oak gall wasps (Cynipidae: Hymenoptera). *Diversity and Distributions*, **8**, 335–349.

Ward RD, Zemlak TS, Innes BH, Last PR & Hebert PDN (2005) DNA bar-coding Australia's fish species. *Philosophical transactions of the Royal Society of London, Series B*, **360**, 1847-1857.

Washburn JO (1984) Mutualism between a Cynipid wasp and ants. *Ecology*, **65**, 654-656.

Wei XX & Wang XQ (2004) Recolonization and radiation in *Larix* (Pinaceae): evidence from nuclear ribosomal DNA paralogues. *Molecular Ecology*, **13**, 3115-3123.

Weld LH (1952). *Cynipoidea (Hym.) 1905-1950 being a Supplement to the Dalla Torre and Kieffer monograph -- the Cynipidae in Das Tierreich, Leiferung 24, 1910 and bringing the systematic literature of the world up to date, including keys to families and subfamilies and list of new generic, specific and variety names.* Ann Arbor, Michigan, Privately Printed. 1-351.

White JM, Ager TA, Adam DP, Leopold EB, Liu G, Jetté H & Schweger CE (1997) An 18 million year record of vegetation and climate change in northwestern Canada and Alaska: tectonic and global climatic correlates. *Plaeogeography, Palaeoclimatology and Palaeoecology*, **130**, 293-306.

Whiting MF, Bradler S & Maxwell T (2003) Loss and recovery of wings in stick insects. *Nature*, **421**, 264-267.

Wiebes-Rijks AA (1976) De Nederlands meeldraad en knopgallen van het *Andricus* Hartig (Hymenoptera, Cynipidae). *Entomologische Berichten, Amsterdam*, **36**, 68-73.

Wiebes-Rijks AA (1979) A character analysis of the species of *Synergus* Hartig, Section II (Mayr, 1872) (Hymenoptera, Cynipidae). *Zoologische Mededelingen*, **53**, 297-321.

Wiebes-Rijks AA (1980) The identity of the gallwasp causing stunted acorns (Hymenoptera: Cynipidae). *Zoologische Mededelingen Leiden*, **53**, 297-321.

Wilcock HR, Hildrew AG & Nichols RA (2001) Genetic differentiation of a European caddisfly: past and present gene flow among fragmented larval habitats. *Molecular Ecology*, **10**, 1821-1834

Wilson DJ & McVean G (2006) Estimating diversifying selection and functional constraint in the presence of recombination. *Genetics*, **172**, 1411-1425.

Wurzell B (2000) The ram's horn gall of *Andricus aries* spreading around north London. *Cecidology*, **15**, 131–134.

Yaltirik F (ed.) (1982) *Fagaceae. Flora of Turkey and the eastern Aegean Islands.* Edinburgh University Press. Edinburgh, UK.

Yang Z & Rannala B (1997) Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. *Molecular Biology and Evolution*, **14**, 717-724.

Yang Z (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *Journal of Molecular Evolution*, **39**, 306-314.

Yang Z (1996) Among-site rate variation and its impact on phylogenetic analyses. *Trends in Ecology and Evolution*, **11**, 367-372.

Yoder AD & Yang ZH (2000) Estimation of primate speciation dates using local molecular clocks. *Molecular Biology and Evolution*, **17**, 1081-1090.

Zheng YQ & Ennos RA (1999). Genetic variability and structure of natural and domesticated populations of Caribbean pine (*Pinus caribaea* Morelet). *Theoretical and Applied Genetics*, **98**, 765–771.

Zink RM, Drovetski SV, Questiau S, Fadeev IV, Nesterov EV, Westberg MC & Rohwer S (2003) Resent evolutionary history of the bluethroat (*Luscinia svecica*) across Eurasia. *Molecular Ecology*, **12**, 3069-3075.

Zohary M (1966) *Flora Palestina*, Vol.1. The Israel Academy of Sciences and Humanities, Goldberg Press, Jerusalem.

Appendix 1

References for studies in Chapter 3 in which geographic origin was unresolved. Studies are numbered as in Table 3.1.

1. Schneller J, Holderegger R, Gugerli F, Eichenberger K & Lutz E (1998) Patterns of genetic variation detected by RAPDs suggest a single origin with subsequent mutations and long-distance dispersal in the apomictic fern *Dryopteris remota* (Dryopteridaceae). *American Journal of Botany*, **85**, 1038-1042.

2. Trewick SA, Morgan-Richards M, Russell SJ, Henderson S, Rumsey FJ, Pinter I, Barrett JA, Gibby M & Vogel JC (2002) Polyploidy, phylogeography and Pleistocene refugia of the rockfern *Asplenium ceterach*: evidence from chloroplast DNA. *Molecular Ecology*, **11**, 2003-2012.

3. Fjellheim S, Rognli OA, Fosnes K & Brochmann C (2006) Phylogeographical history of the widespread meadow fescue (*Festuca pratensis* Huds.) inferred from chloroplast DNA sequences. *Journal of Biogeography*, **33**, 1470-1478.

4. Sharbel TF, Haubold B & Mitchell-Olds T (2000) Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. *Molecular Ecology*, **9**, 2109-2118.

5. King RA & Ferris C (2000) Chloroplast DNA and nuclear DNA variation in the sympatric alder species, *Alnus cordata* (Lois.) Duby and *A. glutinosa* (L.) Gaertn. *Biological Journal of the Linnean Society*, **70**, 147-160.

6. Hampe A, Arroyo J, Jordano P & Petit RJ (2003) Rangewide phylogeography of a bird-dispersed Eurasian shrub: contrasting Mediterranean and temperate glacial refugia. *Molecular Ecology*, **12**, 3415-3426.

7. Palme AE, Su Q, Rautenberg A, Manni F & Lascoux M (2003) Postglacial recolonization and cpDNA variation of silver birch, *Betula pendula*. *Molecular Ecology*, **12**, 201-212.

8. Fineschi S, Taurchini D, Villani F & Vendramin GG (2000) Chloroplast DNA polymorphism reveals little geographical structure in *Castanea sativa* Mill. (Fagaceae) throughout southern European countries. *Molecular Ecology*, **9**, 1495-1503.

9. Demesure B, Comps B & Petit RJ (1996) Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L) in Europe. *Evolution*, **50**, 2515-2520.

10. Magri D, Vendramin GG, Comps B, Dupanloup I, Geburek T, Gomory D, Latalowa M, Litt T, Paule L, Roure JM, Tantau I, van der Knaap WO, Petit RJ & de Beaulieu JL (2006) A new scenario for the Quaternary history of European

beech populations: palaeobotanical evidence and genetic consequences. *New Phytologist*, **171**, 199-221.

11. DumolinLapegue S, Demesure B, Fineschi S, LeCorre V & Petit RJ (1997) Phylogeographic structure of white oaks throughout the European continent. *Genetics*, **146**, 1475-1487.

12. Rendell S, & Ennos RA (2003) Chloroplast DNA diversity of the dioecious European tree *Ilex aquifolium* L. (English holly). *Molecular Ecology*, **12**, 2681-2688.

13. Stone G, Atkinson R, Rokas A, Csóka G, Nieves-Aldrey JL (2001) Differential success in northwards range expansion between ecotypes of the marble gallwasp *Andricus kollari*: a tale of two lifecycles. *Molecular Ecology*, **10**, 761-778.

14. Horn A, Roux-Morabito G, Lieutier F & Kerdelhue C (2006) Phylogeographic structure and past history of the circum-Mediterranean species *Tomicus destruens* Woll. (Coleoptera : Scolytinae). *Molecular Ecology*, **15**, 1603-1615.

15. Stauffer C, Lakatos F & Hewitt GM (1999) Phylogeography and postglacial colonization routes of *Ips typographus* L. (Coleoptera, Scolytidae). *Molecular Ecology*, **8**, 763-773.

16. Rokas A, Atkinson RJ, Brown G, West SA & Stone GN (2001) Understanding patterns of genetic diversity in the oak gall wasp *Biorhiza pallida*: demographic history or a *Wolbachia* selective sweep? *Heredity*, **87**, 294–305.

17. Hayward A & Stone GN (2006) Comparative phylogeography across two trophic levels: the oak gall wasp *Andricus kollari* and its chalcid parasitoid *Megastigmus stigmatizans*. *Molecular Ecology*, **15**, 479-489.

18. Forister ML, Fordyce JA & Shapiro AM (2004) Geological barriers and restricted gene flow in the holarctic skipper *Hesperia comma* (Hesperiidae). *Molecular Ecology*, **13**, 3489-3499.

19. Goropashnaya AV, Fedorov VB, Seifert B & Pamilo P (2004) Limited phylogeographical structure across Eurasia in two red wood ant species *Formica pratensis* and *F. lugubris* (Hymenoptera, Formicidae). *Molecular Ecology*, **13**, 1849-1858.

20. Vandewoestijne S, Baguette M, Brakefield PM & Saccheri IJ (2004) Phylogeography of *Aglais urticae* (Lepidoptera) based on DNA sequences of the mitochondrial COI gene and control region. *Molecular phylogenetics and Evolution*, **31**, 630-641.

21. Durand JD, Persat H & Bouvet Y (1999) Phylogeography and postglacial dispersion of the chub (*Leuciscus cephalus*) in Europe. *Molecular Ecology*, **8**, 989-997.

22. Nesbo CL, Fossheim T, Vollestad LA & Jakobsen KS (1999) Genetic divergence and phylogeographic relationships among European perch (Perca fluviatilis) populations reflect glacial refugia and postglacial colonization. *Molecular Ecology*, **8**, 1387-1404.

23. Palo JU, Schmeller DS, Laurila A, Primmer CR, Kuzmin SL & Merila J (2004) High degree of population subdivision in a widespread amphibian. *Molecular Ecology*, **13**, 2631-2644.

24. Szymura JM, Uzzell T & Spolsky C (2000) Mitochondrial DNA variation in the hybridizing fire-bellied toads, *Bombina bombina* and *B. variegate*. *Molecular Ecology*, **9**, 891-899.

25. Kvist L, Ruokonen M, Lumme J & Orell M (1999) Different population structures in northern and southern populations of the European blue tit. *Journal of Evolutionary Biology*, **12**, 798-805.

26. Pitra C, Lieckfeldt D & Alonso JC (2000) Population subdivision in Europe's great bustard inferred from mitochondrial and nuclear DNA sequence variation. *Molecular Ecology*, **9**, 1165-1170.

27. Liukkonen-Anttila T, Uimaniemi L, Orell M & Lumme J (2002) Mitochondrial DNA variation and the phylogeography of the grey partridge (*Perdix perdix*) in Europe: from Pleistocene history to present day populations. *Journal of Evolutionary Biology*, **15**, 971-982.

28. Zink RM, Drovetski SV, Questiau S, Fadeev IV, Nesterov EV, Westberg MC & Rohwer S (2003) Recent evolutionary history of the bluethroat (*Luscinia svecica*) across Eurasia. *Molecular Ecology*, **12**, 3069-3075.

29. Brito PH (2005) The influence of Pleistocene glacial refugia on tawny owl genetic diversity and phylogeography in western Europe. *Molecular Ecology*, **14**, 3077-3094.

30. Lauga B, Cagnon C, D'Amico F, Karama S & Mouches C (2005) Phylogeography of the white-throated dipper *Cinclus cinclus* in Europe. *Journal of Ornithology*, **146**, 257-262.

31. Guillaumet A, Pons JM, Godelle B & Crochet PA (2006) History of the Crested Lark in the Mediterranean region as revealed by mtDNA sequences and morphology. *Molecular Phylogenetics and Evolution*, **39**, 645-656.

32. Questiau S, Gielly L, Clouet M & Taberlet P (1999) Phylogeographical evidence of gene flow among Common Crossbill (*Loxia curvirostra*, Aves, Fringillidae) populations at the continental level. *Heredity*, **83**, 196-205.

33. Zink RM, Rohwer S, Drovetski S, Blackwell-Rago RC & Farrell SL (2002) Holarctic phylogeography and species limits of Three-toed Woodpeckers. *Condor*, **104**, 167-170.

34. Santucci F, Emerson BC & Hewitt GM (1998) Mitochondrial DNA phylogeography of European hedgehogs. *Molecular Ecology*, **7**, 1163-1172.

35. Seddon JM, Santucci F, Reeve NJ & Hewitt GM (2001) DNA footprints of European hedgehogs, *Erinaceus europaeus* and *E. concolor*. Pleistocene refugia, postglacial expansion and colonization routes. *Molecular Ecology*, **10**, 2187-2198.

36. Berggren KT, Ellegren H, Hewitt GM & Seddon JM (2005) Understanding the phylogeographic patterns of European hedgehogs, *Erinaceus concolor* and *E. europaeus* using the MHC. *Heredity*, **95**, 84-90.

37. Wiehler J & Tiedemann R (1998) Phylogeography of the European roe deer *Capreolus capreolus* as revealed by sequence analysis of the mitochondrial Control Region. *Acta Theriologia*, Supp. 5, 187-197.

38. Randi E, Alves PC, Carranza J, Milosevic-Zlatanovic S, Sfougaris A & Mucci N (2004) Phylogeography of roe deer (*Capreolus capreolus*) populations: the effects of historical genetic subdivisions and recent nonequilibrium dynamics. *Molecular Ecology*, **13**, 3071-3083.

39. Lorenzini R & Lovari S (2006) Genetic diversity and phylogeography of the European roe deer: the refuge area theory revisited. *Biological Journal of the Linnean Society*, **88**, 85-100.

40. Davison A, Griffiths HI, Brookes RC, Maran T, Macdonald DW, Sidorovich VE, Kitchener AC, Irizar I, Villate I, Gonzalez-Esteban J, Cena JC, Cena A, Moya I & Minano SP (2000) Mitochondrial DNA and palaeontological evidence for the origins of endangered European mink, *Mustela lutreola*. *Animal Conservation*, **3**, 345-355.

41. Michaux JR, Libois R, Paradis E & Filippucci MG (2004) Phylogeographic history of the yellow-necked fieldmouse (*Apodemus flavicollis*) in Europe and in the Near and Middle East. *Molecular Phylogenetics and Evolution*, **32**, 788-798.

42. Michaux JR, Libois R & Filippucci MG (2005) So close and so different: comparative phylogeography of two small mammal species, the Yellow-necked fieldmouse (*Apodemus flavicollis*) and the Woodmouse (*Apodemus sylvaticus*) in the Western Palearctic region. *Heredity*, **94**, 52-63.

43. Durka W, Babik W, Ducroz JF, Heidecke D, Rosell F, Samjaa R, Saveljev AP, Stubbe A, Ulevicius A & Stubbe M (2005) Mitochondrial phylogeography of the Eurasian beaver *Castor fiber* L. *Molecular Ecology*, **14**, 3843-3856.

44. Kasapidis P, Suchentrunk F, Magoulas A & Kotoulas G (2005) The shaping of mitochondrial DNA phylogeographic patterns of the brown hare (*Lepus europaeus*) under the combined influence of Late Pleistocene climatic fluctuations and anthropogenic translocations. *Molecular Phylogenetics and Evolution*, **34**, 55-66.

45. Neumann K, Michaux JR, Maak S, Jansman HAH, Kayser A, Mundt G & Gattermann R (2005) Genetic spatial structure of European common hamsters (*Cricetus cricetus*) - a result of repeated range expansion and demographic bottlenecks. *Molecular Ecology*, **14**, 1473-1483.

46. Bilgin R, Karatas A, Coraman E, Pandurski I, Papadatou E & Morales JC (2006) Molecular taxonomy and phylogeography of *Miniopterus schreibersii* (Kuhl, 1817) (Chiroptera : Vespertilionidae), in the Eurasian transition. *Biological Journal of the Linnean Society*, **87**, 577-582.

47. Taberlet P & Bouvet J (1994) Mitochondrial-DNA polymorphism, phylogeography, and conservation genetics of the brown bear *Ursus arctos* in Europe. *Proceedings of the Royal Society of London, Series B*, **255**, 195-200.

48. Sommer RS & Benecke N (2005) The recolonization of Europe by brown bears *Ursus arctos* Linnaeus, 1758 after the Last Glacial Maximum. *Mammal Review*, **35**, 156-164.

49. Davison A, Birks JDS, Brookes RC, Messenger JE & Griffiths HI (2001) Mitochondrial phylogeography and population history of pine martens *Martes martes* compared with polecats *Mustela putorius*, **10**, 2479-2488.

50. Michaux JR, Magnanou E, Paradis E, Nieberding C, Libois R (2003) Mitochondrial phylogeography of the Woodmouse (*Apodemus sylvaticus*) in the Western Palearctic region. *Molecular Ecology*, **12**, 685-697.

51. Brunhoff C, Galbreath KE, Fedorov VB, Cook JA & Jaarola M (2003) Holarctic phylogeography of the root vole (*Microtus oeconomus*): implications for late Quaternary biogeography of high latitudes. *Molecular Ecology*, **12**, 957-968.

52. Ferrando A, Ponsa M, Marmi J, Domingo-Roura X (2004) Eurasian otters, *Lutra lutra*, have a dominant mtDNA haplotype from the Iberian Peninsula to Scandinavia. *Journal of Heredity*, **95**, 430-435.

53. Heckel G, Burri R, Fink S, Desmet JF & Excoffier L (2005) Genetic structure and colonization processes in European populations of the common vole, *Microtus arvalis*. *Evolution*, **59**, 2231-2242.

54. Kotlik P, Deffontaine V, Mascheretti S, Zima J, Michaux JR & Searle JB (2006) A northern glacial refugium for bank voles (*Clethrionomys glareolus*). *Proceedings of the National Academy of Science USA*, **103**, 14860-14864.

55. Pope LC, Domingo-Roura X, Erven K & Burke T (2006) Isolation by distance and gene flow in the Eurasian badger (*Meles meles*) at both a local and broad scale. *Molecular Ecology*, **15**, 371-386.

Appendix 2

Locations, sample sizes and haplotypes sampled for each population in Chapter 4. Locations are given in decimal degrees for latitude followed by longitude, and are shown in Figure 4.1. Where multiple copies of a haplotype were sampled from a population, numbers in parentheses after haplotype number indicate the number of individuals sharing that haplotype.

	Population	Country	Location	Sample size	Haplotypes
1	El Escorial	Spain	40.58,-4.13	3	18(2), 32
2	Orusco	Spain	40.28,-3.22	1	18
3	Llerida	Spain	41.61,0.63	1	18
4	Valpiana	Italy	43.02,10.84	1	1
5	Anconella	Italy	43.76,11.30	1	12
6	Abruzzo	Italy	42.49,13.72	2	9,1
7	Cupoli	Italy	42.46,13.83	2	8,9
8	Molise	Italy	41.68,14.54	2	1(2)
9	Gargano	Italy	41.89,16.13	2	10,11
10	Istria	Croatia	45.21,13.89	1	3
11	Köszeg	Hungary	47.39,16.54	3	4,8,20
12	Várpalota	Hungary	47.20,18.15	1	5
13	Tatabánya	Hungary	47.60,18.41	1	1
14	Szoloske	Hungary	47.88,19.01	2	6,7
15	Gödöllö	Hungary	47.61,19.36	1	16
16	Balaton	Hungary	48.10,20.31	1	9
17	Eger	Hungary	47.92,20.38	1	2
18	Plástovce	Slovakia	48.16,18.98	1	5
19	Prespa	Greece	40.77,21.09	1	9
20	Pisoderi	Greece	40.78,21.25	4	5,9,13,15
21	Florina	Greece	40.78,21.41	1	14
22	Komnina	Greece	40.59,21.77	1	17
23	Edessa	Greece	40.80,22.05	1	5
24	Tarsus	Turkey	36.92,34.90	1	23
25	Küllüce	Turkey	38.20,34.60	1	16
26	Çekerek	Turkey	40.07,35.49	1	24
27	Suluova	Turkey	40.84,35.65	3	16(2),21
28	Tokat	Turkey	40.32,36.55	2	16,19
29	Niksar	Turkey	40.59,36.95	3	22,25(2)
30	Deir al Zahari	Lebanon	33.43,35.47	2	43, 44
31	Jezzine	Lebanon	33.54,35.59	1	43
32	Ain Dara	Lebanon	33.78,35.73	1	41
33	Ibrahim River	Lebanon	34.07,35.88	1	42
34	Pyranshahr	Iran	36.69,45.23	4	26(2),27,28
35	Baneh	Iran	35.99,45.90	14	16(6), 29, 30, 31(5), 33
36	Marivan	Iran	35.52,46.17	11	34(2), 35(4), 36, 37, 38, 39, 40

Appendix 3

Sample sites and sample sizes for 13 polymorphic allozyme loci in *Andricus kollari* for Chapter 5. Locations are given in decimal degrees latitude followed by longitude, with a negative sign for western coordinates. Data for the other sites are available on line at http://www.blackwellpublishing.com/products/journals/suppmat /mec/mec1211/mec1211sm.htm

Site	Site Name	Country	Location	Allozyme sample size
1	Truro	UK	50.27,-5.05	9
2	Cardinham	UK	50.48,-4.65	32
3	Bovey Tracey	UK	50.58,-3.68	8
4	Sherbourne	UK	50.95,-2.50	9
5	Burley	UK	50.81,-1.70	44
6	Puttenham	UK	51.20,-0.72	42
7	Ashford	UK	51.13,0.88	12
8	Oxford	UK	51.75,-1.25	33
9	Romney	UK	52.33,-1.95	22
10	Swansea	UK	51.63,-3.97	26
11	Muckross	IRE	51.22,-10.8	29
12	Fota	IRE	51.54,-8.21	30
13	Isle of Man	UK	54.15,-4.48	41
14	Knutsford	UK	53.30,-2.36	27
15	Stone	UK	52.90,-2.13	11
16	Chatsworth	UK	53.20,-1.66	42
17	Rufford Park	UK	53.16,-1.04	44
18	Skares	UK	55.42,-4.32	13
19	Erskine	UK	55.91,-4.46	44
20	Kyle of Lochalsh	UK	57.27,-5.71	14
21	Beauly	UK	57.40,-4.46	41
22	Dunrobin	UK	57.98,-3.94	48
23	Treveles	SPA	36.97,-3.27	38
24	Yegen	SPA	36.96,-3.13	26
25	Laroles	SPA	37.00,-3.00	19
26	Navacerrada	SPA	40.73,-4.02	44
27	Salamanca	SPA	40.96,-5.67	35
28	Gudiña	SPA	42.07,-7.13	67
29	Quintanilla	SPA	42.15,-6.27	22
30	Madrid	SPA	40.65,-4.15	16
31	Potes	SPA	43.15,-4.62	17
32	Bayonne	FRA	43.50,-1.47	25
33	Amou	FRA	43.58,-0.75	44
34	Mugron	FRA	43.75,-0.75	25
35	Pau	FRA	43.30,-0.38	24
36	Auch	FRA	43.65,0.56	18
37	Toulouse	FRA	43.60,1.41	24
38	Agen	FRA	44.20,0.62	30
39	Bordeaux	FRA	44.85,-0.56	28
40	La Rochelle	FRA	46.17,-1.16	44

Site	Site Name	Country	Location	Allozyme sample size
41	Nantes	FRA	47.22,-1.57	17
42	Rennes	FRA	48.12,-1.68	19
43	St. Malo	FRA	48.65,-2.02	22
44	Mortain	FRA	48.65,0.93	44
45	Coutances	FRA	49.05,-1.45	43
46	Saumur	FRA	47.30,-0.07	32
47	Nouvion	FRA	50.20,1.77	12
48	Forêt de Bercé	FRA	47.78,0.38	44
49	Crécy	FRA	50.23,1.83	44
50	Torhout	BEL	51.05,3.10	14
51	Amsterdam	NL	52.35,4.86	15
52	Utrecht	NL	52.10,5.10	19
53	Krefeld	GER	48.90,9.20	10
54	Köln	GER	50.95,6.90	24
55	Ruffeno	IT	44.20,10.93	60
56	Greve	IT	43.58,11.20	17
57	Sopron	HUN	47.68,16.57	35
58	Gödöllö	HUN	47.58,19.33	16
59	Mátrafüred	HUN	47.84,19.95	44
60	Fehérgyarmat	HUN	47.98,22.50	44
61	Hortobágy	HUN	47.58,21.15	44
62	Berettyóújfalu	HUN	47.22,21.52	44
63	Szeghalom	HUN	47.02,21.17	32
64	Gyula	HUN	46.63,21.27	19
65	Öcsöd	HUN	46.90,20.38	37
66	Sellje	HUN	45.85,17.83	44
67	Nagykanizsa	HUN	46.45,16.98	44
68	Zalaegerszeg	HUN	46.83,16.82	44
69	Antalya	TUR	36.90,30.70	42

Appendix 4

Allele frequencies by site for UK and Irish *Andricus kollari* for Chapter 5. Site details are provided by site number in Appendix 3.

	Site									
Locus	1 Truro	2 Cardin	3 B.Trac	4 Sher	5 Burl	6 Putt	7 Ash	8 Oxf	9 Rom	10 Swa
GPD1 1	0	0	0	0	0	0	0.042	0	0	0
GPD1 4	1	1	1	1	1	1	0.958	1	1	1
GPD2 2	1	1	1	1	1	0.988	1	1	1	1
GPD2 3	0	0	0	0	0	0	0	0	0	0
GPD2 4	0	0	0	0	0	0.012	0	0	0	0
GOT-s 1	0	0.109	0	0	0.432	0.012	0	0	0	0.288
GOT-s 2	1	0.875	1	1	0.568	0.988	1	1	1	0.712
GOT-s 3	0	0.016	0	0	0	0	0	0	0	0
GOT-m 1	0	0.047	0.313	0.056	0.080	0.036	0	0.121	0.045	0.019
GOT-m 2	0.722	0.578	0.188	0.333	0.352	0.405	0.125	0.364	0.250	0.423
GOT-m 3	0.111	0.109	0.375	0.167	0.341	0.333	0.500	0.348	0.455	0.269
GOT-m 4	0.167	0.266	0.125	0.444	0.227	0.226	0.375	0.167	0.250	0.288
GPI 1	0	0	0	0	0.034	0	0.083	0	0	0
GPI 2	0	0.109	0.250	0.056	0.091	0.226	0.042	0.197	0.136	0.038
GPI 3	1	0.813	0.750	0.778	0.875	0.762	0.875	0.803	0.864	0.808
GPI 4	0	0.078	0	0.167	0	0.012	0	0	0	0.154
HK 1	0	0	0	0.056	0	0	0.083	0.015	0	0
HK 2	1	1	1	0.944	1	1	0.917	0.985	1	1
HK 3	0	0	0	0	0	0	0	0	0	0
PEB-b 1	0	0	0	0	0	0	0	0.015	0	0
PEB-b 2	0.333	0.5	0.250	0.222	0.318	0.131	0.292	0.136	0.364	0.635
PEB-b 3	0.056	0.219	0.500	0.722	0.352	0.524	0.500	0.591	0.364	0.154
PEB-b 4	0.167	0.125	0.250	0.056	0.330	0.333	0.208	0.258	0.273	0.038
PEB-b 5	0.444	0.156	0	0	0	0	0	0	0	0.173
PEB-b 6	0	0	0	0	0	0.012	0	0	0	0
PGM 1	0	0.031	0	0	0.011	0.060	0	0.061	0.023	0
PGM 2	0.167	0.391	0.813	0.111	0.648	0.679	0.625	0.788	0.750	0.077
PGM 3	0.444	0.563	0.188	0.778	0.261	0.167	0.167	0.121	0.182	0.846
PGM 4	0.389	0.016	0	0.111	0.080	0.095	0.208	0.030	0.045	0.077
AK 1	0	0.297	0	0.222	0	0	0	0	0	0.077
AK 2	0	0.078	0	0	0	0.012	0	0	0	0.058
AK 3	1	0.531	1	0.778	1	0.988	1	1	1	0.865
AK 4	0	0.094	0	0	0	0	0	0	0	0
6PGD 1	0	0.063	0	0	0	0	0	0	0	0
6PGD 2	1	0.703	0.875	0.5	0.731	0.679	0.375	0.742	1	0.269
6PGD 3	0	0	0	0	0	0	0	0	0	0
6PGD 4	0	0.188	0.063	0.111	0.115	0.143	0.375	0.152	0	0.308
6PGD 5	0	0	0	0.333	0.026	0.024	0	0	0	0.385
6PGD 6	0	0.016	0	0.056	0	0.036	0.167	0.015	0	0.019
6PGD 7	0	0	0.063	0	0.064	0.048	0.042	0.015	0	0
6PGD 8	0	0.031	0	0	0.064	0.071	0.042	0.076	0	0.019

	Site									
Locus	11 Muck	12 Fota	13 IOM	14 Knut	15 Ston	16 Chat	17 Ruff	18 Skar	19 Ersk	20 Kyl
GPD1 1	0	0	0	0	0	0	0	0	0	0
GPD1 4	1	1	1	1	1	1	1	1	1	1
GPD2 2	1	1	1	1	1	1	1	1	1	1
GPD2 3	0	0	0	0	0	0	0	0	0	0
GPD2 4	0	0	0	0	0	0	0	0	0	0
GOT-s 1	0	0	0	0.241	0.182	0.286	0.261	0	0.045	0
GOT-s 2	1	1	1	0.741	0.818	0.702	0.739	1	0.955	1
GOT-s 3	0	0	0	0.019	0	0.012	0	0	0	0
GOT-m 1	0.017	0	0	0.204	0	0	0.011	0	0	0.036
GOT-m 2	0.552	0.317	0.220	0.444	0.682	0.512	0.318	0.423	0.523	0.143
GOT-m 3	0.190	0.500	0.390	0.037	0.045	0.012	0.045	0.538	0.068	0.464
GOT-m 4	0.241	0.183	0.390	0.315	0.273	0.476	0.625	0.038	0.409	0.357
GPI 1	0.017	0	0	0	0.045	0	0	0	0	0.071
GPI 2	0.190	0.100	0.037	0.296	0.273	0.167	0.136	0	0.011	0
GPI 3	0.793	0.900	0.963	0.537	0.545	0.679	0.568	1	0.955	0.929
GPI 4	0	0	0	0.167	0.136	0.155	0.295	0	0.034	0
HK 1	0	0	0	0	0	0	0	0	0	0
HK 2	1	1	1	1	1	1	1	1	1	1
HK 3	0	0	0	0	0	0	0	0	0	0
PEB-b 1	0.034	0	0	0	0	0	0	0	0	0
PEB-b 2	0.276	0.333	0.192	0.556	0.750	0.702	0.807	0.115	0.116	0.321
PEB-b 3	0.362	0.433	0.577	0.037	0.050	0.170	0.023	0.654	0.767	0.429
PEB-b 4	0.328	0.233	0.231	0.037	0.050	0.012	0	0.231	0.058	0.250
PEB-b 5	0	0	0	0.370	0.150	0.179	0.170	0	0.058	0
PEB-b 6	0	0	0	0	0	0	0	0	0	0
PGM 1	0.052	0.083	0	0.056	0	0	0	0	0.013	0
PGM 2	0.707	0.800	0.805	0.056	0.136	0.095	1	0.808	0.487	0.929
PGM 3	0.241	0.117	0.122	0.870	0.727	0.905	0	0	0.436	0.036
PGM 4	0	0	0.073	0.019	0.136	0	0	0.192	0.064	0.036
AK 1	0.017	0	0	0.400	0.182	0.363	0.239	0	0.095	0
AK 2	0	0	0	0.080	0.136	0.075	0.080	0	0	0
AK 3	0.983	1	1	0.520	0.682	0.563	0.682	1	0.905	1
AK 4	0	0	0	0	0	0	0	0	0	0
6PGD 1	0	0.017	0	0	0	0	0.012	0	0	0
6PGD 2	0.724	0.833	0.730	0.444	0.636	0.387	0.268	1	0.674	0.864
6PGD 3	0	0	0	0	0	0	0	0	0	0
6PGD 4	0.172	0.117	0.149	0.111	0.364	0.075	0.073	0	0.070	0.136
6PGD 5	0	0	0	0.407	0	0.512	0.646	0	0.244	0
6PGD 6	0	0	0.014	0.019	0	0	0	0	0	0
6PGD 7	0.034	0.033	0	0	0	0	0	0	0	0
6PGD 8	0.069	0	0.108	0.019	0	0.025	0	0	0.012	0

Locus	Site 21 Beau	Site 22 Dunr
GPD1 1	0	0
GPD1 4	1	1
GPD2 2	1	0.979
GPD2 3	0	0.021
GPD2 4	0	0
GOT-s 1	0	0.021
GOT-s 2	1	0.979
GOT-s 3	0	0
GOT-m 1	0.012	0.021
GOT-m 2	0.341	0.191
GOT-m 3	0.488	0.500
GOT-m 4	0.159	0.287
GPI 1	0	0.011
GPI 2	0.128	0.120
GPI 3	0.872	0.848
GPI 4	0	0.022
HK 1	0	0
HK 2	1	0.963
HK 3	0	0.037
PEB-b 1	0	0
PEB-b 2	0.098	0.167
PEB-b 3	0.415	0.456
PEB-b 4	0.488	0.378
PEB-b 5	0	0
PEB-b 6	0	0
PGM 1	0.024	0.233
PGM 2	0.732	0.544
PGM 3	0.159	0.189
PGM 4	0.085	0.033
AK 1	0	0
AK 2	0	0
AK 3	1	1
AK 4	0	0
6PGD 1	0	0
6PGD 2	0.950	0.894
6PGD 3	0	0
6PGD 4	0	0.053
6PGD 5	0	0
6PGD 6	0	0
6PGD 7	0	0
6PGD 8	0.050	0.053

Appendix 5

Sample sites, sample sizes and haplotype numbers (with numbers of copies when >1 in brackets) for mitochondrial *cytb* and nuclear 28S D2 sequences for Chapter 5. *Cytb* haplotype numbers are consistent with Hayward & Stone (2006), and numbers of contributing sequences previously published in Stone *et al.* (2001) (total 27 sequences, Genbank accession numbers AF242739-AF242762 and AF242764-AF242766) and Hayward & Stone (2006) (total 27 sequences, Genbank accession numbers DQ925335-DQ925361) are given in brackets in the *cytb* sample size column. Locations are given in decimal degrees latitude followed by longitude, with a negative sign for western coordinates. Site numbers refer to the map in Figure 5.3.

Site	Site name	Country	Location	Cytb sample size	Cytb haplotypes	D2 haplotypes
33	Amou	FRA	43.58,-0.75	4(2)	23(2), 24, 36	1/2, 2
32	Bayonne	FRA	43.50,-1.47	6(3)	23(3), 26(2), 40	1/2, 2
39	Bordeaux	FRA	44.85,-0.56	1	44	1
-	Ceret	FRA	42.49,2.74	3(3)	36(3)	
49	Crécy	FRA	50.23,1.83	5(5)	8, 15(2), 17, 19	
40	La Rochelle	FRA	46.17,-1.16	2(2)	3, 10	
	Montpellier	FRA	43.60,3.85	1(1)	27	
44	Mortain	FRA	48.65,0.93	6(2)	20, 21, 54, 63, 64, 65	1(3)
34	Mugron	FRA	43.75,-0.75	2(2)	22, 23	
41	Nantes	FRA	47.22,-1.57	6(1)	18, 52, 53, 54, 55, 82	1
35	Pau	FRA	43.30,-0.38	2(2)	23, 30	
-	Pierroton	FRA	43.37,5.31	3(3)	23, 26, 32	
42	Rennes	FRA	48.12,-1.68	3(3)	2, 9, 11	
43	St. Malo	FRA	48.65,-2.02	5(2)	7(3), 21, 55	1(2)
48	Vannes	FRA	47.68,-2.75	3(3)	1, 4, 12	
54	Köln	GER	50.95,6.90	3	18, 46, 51	1(3)
60	Fehérgyarmat	HUN	47.98,22.50	2	66, 67	1(2)
58	Gödöllö	HUN	47.58,19.33	4(4)	5, 7(2), 13	
64	Gyula	HUN	46.63,21.27	2	68, 81	
57	Sopron	HUN	47.68,16.57	4	5(2), 56, 62	1(4)
68	Zalaegerszeg	HUN	46.83,16.82	3	7, 56, 57	
-	Az	IRAN	0,0	1	7	1
-	Piran Shahr	IRAN	36.70,45.13	2	94, 95	1
11	Muckross	IRE	51.22,-10.86	3	77, 78, 79	1(2)
-	Etna	IT	37.76,14.99	1	5	
-	Furore	IT	40.62,14.54	2	41, 95	1
56	Greve	IT	43.58,11.20	3(1)	6, 48, 58	1(2)
55	Ruffeno	IT	44.20,10.93	3(2)	15, 16, 45	
-	Sangin	IT	0,0	2	98(2)	
-	Florence	IT	0,0	5	6, 69, 70, 71, 72	1(3)

Site	Site name	Country	Location	Cytb sample size	Cytb haplotypes	D2 haplotypes
51	Amsterdam	NL	52.35,4.86	3	49, 50, 60	1(3)
52	Utrecht	NL	52.10,5.10	3	59, 60, 61	
-	Lagarelhos	PORT	41.10,-8.15	1	34	
-	El Escorial	SPA	40.58,-4.15	2(2)	31, 33	
31	Potes	SPA	43.15,-4.62	3(2)	23, 28, 42	
29	Quintanilla	SPA	42.15,-6.27	4(1)	25(3), 43	
-	Rubi	SPA	41.47,2.25	2(2)	32, 35	
27	Salamanca	SPA	40.96,-5.67	3(1)	23(2), 26	1/2, 2
23	Treveles	SPA	36.97,-3.27	3	37, 38, 39	1/2, 2
69	Antalya	TUR	36.90,30.70	5(2)	14(4), 47	1(2)
-	Bolu	TUR	40.74,31.61	7	7, 56, 59, 92, 93, 96(2), 97	1(4)
	Sariz (Kayseri)	TUR	38.67,36.32	4	105, 106, 107, 108	
	Gevas (Bitlis)	TUR	38.47,42.52	3	102, 103, 104	
	Asagi Kolbasi (Mus)	TUR	38.55,42.09	3	99, 100, 101	
2	Cardinham	UK	50.48,-4.65	7	20, 46, 75, 76, 83, 88, 90	1(6)
16	Chatsworth	UK	53.20,-1.66	1	85	1
22	Dunrobin	UK	57.98,-3.94	3	46, 73, 80	1
19	Erskine	UK	55.91,-4.46	1	18	1
14	Knutsford	UK	53.30,-2.36	2	87, 91	
8	Oxford	UK	51.75,-1.25	3	61, 74, 84	1
17	Rufford Park	UK	53.16,-1.04	1	86	1
10	Swansea	UK	51.63,-3.97	1	89	1/2

Appendix 6

Summary of identified species and sample location details for each of the 174 *cytb* and two 28SD2 haplotypes for Chapter 6. Locations are given in decimal degrees latitude followed by longitude. Preliminary sequence analysis revealed that *cytb* haplotypes 11, 27 and 28 were not members of the *A. quercuscalicis* clade so these sequences (and the associated 28SD2 haplotype 2) were excluded from subsequent analyses.

cytb	D2	Species	Country	Site	Location	Accession
1	1	*A. caputmedusae*	Croatia	Istria	45.26,13.94	
1		*A. caputmedusae*	Croatia	Istria	45.26,13.94	
1		*A. caputmedusae*	Czech Republic	Valtice	48.74,16.76	AJ228456
1	1	*A. caputmedusae*	Greece	North of Florina	40.78,21.37	
1		*A. caputmedusae*	Greece	North of Florina	40.78,21.37	
1		*A. caputmedusae*	Hungary	Sirok	47.93,20.21	
1	1	*A. caputmedusae*	Hungary	Szentkut	48.00,20.00	
1	1	*A. caputmedusae*	Hungary	Varpolota	47.20,18.15	
1	1	*A. caputmedusae*	Hungary	Varpolota	47.20,18.15	
1	1	*A. caputmedusae*	Italy	Foresta Umbra	41.92,15.60	
1	1	*A. caputmedusae*	Italy	Gildone	41.50,14.67	
1		*A. caputmedusae*	Italy	Gildone	41.50,14.67	
1	1	*A. caputmedusae*	Italy	Monte Vulture	40.64,15.80	
1	1	*A. dentimitratus*	Hungary	Koszeg	47.39,16.54	
1		*A. dentimitratus*	Italy	Foresta Umbra	41.92,15.60	
1	1	*A. dentimitratus*	Italy	Foresta Umbra	41.92,15.60	
1	1	*A. dentimitratus*	Italy	Massa Marittima	43.06,10.89	
1	1	*A. dentimitratus*	Italy	Valpiana	43.02,10.84	
2		*A. caputmedusae*	Croatia	Istria	45.26,13.94	
2	1	*A. caputmedusae*	Hungary	Szenkut	47.98,19.78	
2	1	*A. caputmedusae*	Hungary	Varpolota	47.20,18.15	
2		*A. dentimitratus*	Hungary	Koszeg	47.39,16.54	
2	1	*A. dentimitratus*	Hungary	Koszeg	47.39,16.54	
3		*A. caputmedusae*	Hungary	Matrafured	47.85,19.99	
4		*A. caputmedusae*	Hungary	Szenkut	47.98,19.78	
5		*A. caputmedusae*	Hungary	Szenkut	47.98,19.78	
6	1	*A. caputmedusae*	Italy	Monte Sant Angelo	41.92,15.60	
7	1	*A. caputmedusae*	Italy	Monte Sant Angelo	41.92,15.60	
8	1	*A. caputmedusae*	Italy	Gargano	41.92,15.60	
8		*A. quercustozae*	Croatia	Istria	45.26,13.94	
8		*A. quercustozae*	Croatia	Istria	45.26,13.94	
8		*A. quercustozae*	France	Aire de Querane	45.37,-0.97	AY157279
8		*A. quercustozae*	Hungary	Godollo	47.60,19.33	AY157279
8		*A. quercustozae*	Hungary	Sopron	47.68,16.59	AY157279
8		*A. quercustozae*	Hungary	Szeghalom	47.23,16.70	AY157279
8		*A. quercustozae*	Italy	Felitto	40.37,15.25	AY157279
8		*A. quercustozae*	Italy	Greve in Chianti	43.58,11.32	AY157279
8		*A. quercustozae*	Italy	Poppi	43.72,11.77	AY157279
8		*A. quercustozae*	Italy	Radicofani	42.09,11.77	AY157279
8		*A. quercustozae*	Italy	Rieti	42.40,12.85	AY157279
9	1	*A. caputmedusae*	Italy	Monte Vulture	40.64,15.80	AF539553
10		*A. caputmedusae*	Italy	Pedemonte	45.53,11.00	
11	2	*A. caputmedusae*	Italy	San Venanzo	42.87,12.27	
12	1	*A. caputmedusae*	Croatia	Istria	45.26,13.94	
13		*A. caputmedusae*	Croatia	Istria	45.26,13.94	
14	1	*A. caputmedusae*	Turkey	Hadim	38.07,29.60	

cytb	D2	Species	Country	Site	Location	Accession
14	1	*A. caputmedusae*	Turkey	Kulluce	38.20,34.60	
14	1	*A. caputmedusae*	Turkey	Suluova	40.84,35.65	
15		*A. caputmedusae*	Turkey	Kirazoglu	39.23,32.44	
15		*A. caputmedusae*	Turkey	Kirazoglu	39.23,32.44	
15	1	*A. caputmedusae*	Turkey	Kulluce	38.20,34.60	AF539554
16	1	*A. caputmedusae*	Turkey	Kulluce	38.20,34.60	
17	1	*A. caputmedusae*	Turkey	Kulluce	38.20,34.60	
18	1	*A. caputmedusae*	Turkey	Suluova	40.84,35.65	
18	1	*A. caputmedusae*	Turkey	Suluova	40.84,35.65	
19	1	*A. caputmedusae*	Turkey	Tokat	40.32,36.55	
20	1	*A. caputmedusae*	Turkey	Tokat	40.32,36.55	
21		*A. caputmedusae*	Iran	Golestan	33.49,48.35	
21		*A. caputmedusae*	Iran	Golestan	33.49,48.35	
22		*A. caputmedusae*	Iran	Golestan	33.49,48.35	
23		*A. caputmedusae*	Hungary	Sirok	47.93,20.20	
24	1	*A. caputmedusae*	Greece	Lake Prespa	40.76,21.15	
25	1	*A. caputmedusae*	Hungary	Sirok	47.93,20.20	
26	1	*A. caputmedusae*	Hungary	Szentkut	48.00,20.00	
27		*A. caputmedusae*	Iran	Golestan	33.49,48.35	
28		*A. caputmedusae*	Iran	Golestan	33.49,48.35	
29		*A. caputmedusae*	Iran	Golestan	33.49,48.35	
30		*A. caputmedusae*	Iran	Golestan	33.49,48.35	
31		*A. caputmedusae*	Iran	Golestan	33.49,48.35	
32	1	*A. caputmedusae*	Italy	Assergi	42.41,13.50	
33	1	*A. caputmedusae*	Italy	Capalbio Malhierba	42.45,11.42	
33	1	*A. caputmedusae*	Italy	Jelsi	41.50,14.80	
33	1	*A. caputmedusae*	Italy	Monzuno	44.28,11.27	
34	1	*A. caputmedusae*	Italy	Monte Sant Angelo	47.70,16.00	
35	1	*A. caputmedusae*	Italy	Monte Vulture	41.50,15.00	
36		*A. quercuscalicis*	Hungary	Godollo	47.60,19.36	
36		*A. quercuscalicis*	Netherlands	Wageningen	51.96,5.660	
36		*A. quercuscalicis*	United Kingdom	A5/M42	51.75,-1.26	AJ228459
36		*A. quercuscalicis*	United Kingdom	Cambridge	52.20,0.120	
36		*A. quercuscalicis*	United Kingdom	Oxford univ parks	51.75,-1.25	
36		*A. quercuscalicis*	United Kingdom	Rothampstead	51.80,-0.35	
36		*A. quercuscalicis*	United Kingdom	Steventon	51.62,-1.33	
36		*A. quercuscalicis*	United Kingdom	Thringstone	52.75,-1.37	
37		*A. quercuscalicis*	Netherlands	Wageningen	51.96,5.660	
38		*A. quercuscalicis*	United Kingdom	Thringstone	52.75,-1.37	
39		*A. quercuscalicis*	United Kingdom	Thringstone	52.75,-1.37	
40		*A. quercuscalicis*	United Kingdom	Rothampstead	52.80,-0.36	
41		*A. quercuscalicis*	France	Crécy	50.25,1.880	
41		*A. quercuscalicis*	Ireland	Glasnevin	53.38,-6.26	
41		*A. quercuscalicis*	United Kingdom	Knightwood oak	50.86,-1.58	
41		*A. quercuscalicis*	United Kingdom	Oxford univ parks	51.75,-1.25	
41		*A. quercuscalicis*	United Kingdom	Steventon	51.62,-1.33	
41	1	*A. quercuscalicis*	United Kingdom	Tiverton	50.86,-3.25	
42		*A. quercuscalicis*	United Kingdom	Oxford univ parks	51.75,-1.25	
43	1	*A. quercuscalicis*	United Kingdom	Knightwood oak	50.86,-1.58	
44		*A. quercuscalicis*	United Kingdom	Oxford univ parks	51.75,-1.25	
45		*A. quercuscalicis*	United Kingdom	Oxford univ parks	51.75,-1.25	
46		*A. quercuscalicis*	United Kingdom	Knightwood oak	50.86,-1.58	
47		*A. quercuscalicis*	Netherlands	Wageningen	51.96,5.660	
49		*A. quercuscalicis*	United Kingdom	Rothampstead	53.80,-0.37	
51		*A. quercuscalicis*	United Kingdom	Steventon	51.62,-1.33	
52		*A. quercuscalicis*	United Kingdom	Steventon	51.62,-1.33	
53		*A. quercuscalicis*	United Kingdom	Thringstone	52.75,-1.37	
54		*A. quercuscalicis*	United Kingdom	Thringstone	52.75,-1.37	

cytb	D2	Species	Country	Site	Location	Accession
55		*A. dentimitratus*	Slovenia	Bled	46.37,14.11	
55		*A. quercuscalicis*	Austria	Rosenau	48.00,14.73	
55		*A. quercuscalicis*	Germany	Munich	48.15,11.58	
55		*A. quercuscalicis*	Hungary	Godollo	47.60,19.36	
56		*A. quercuscalicis*	Hungary	Sopron	16.59,47.68	
57		*A. quercuscalicis*	Hungary	Miskolc	48.10,20.78	
58		*A. quercuscalicis*	Hungary	Miskolc	48.10,20.78	
59		*A. quercuscalicis*	Hungary	Miskolc	48.10,20.78	
60	1	*A. quercuscalicis*	Hungary	Miskolc	48.10,20.78	
61	1	*A. quercuscalicis*	Hungary	Miskolc	48.10,20.78	
62		*A. quercuscalicis*	Romania	Sibiu	45.80,24.15	
63		*A. quercuscalicis*	Italy	Asti	44.90,8.200	
64		*A. quercuscalicis*	Italy	Asti	44.90,8.200	
65		*A. quercuscalicis*	Belgium	Bruges	51.21,3.230	
66		*A. quercuscalicis*	Belgium	Bruges	51.21,3.230	
67		*A. quercuscalicis*	France	Crécy	50.25,1.880	
68		*A. quercuscalicis*	France	Crécy	50.25,1.880	
69		*A. quercuscalicis*	Germany	Munich	48.15,11.58	
70		*A. quercuscalicis*	Italy	Asti	44.90,8.200	
71	1	*A. quercuscalicis*	Italy	Asti	44.90,8.200	
72		*A. quercuscalicis*	Germany	Ludwigsburg	50.50,12.42	
72		*A. quercuscalicis*	Slovenia	Ljubljana	46.05,14.51	
73		*A. quercuscalicis*	Germany	Ludwigsburg	50.50,12.42	
74		*A. quercuscalicis*	Germany	Ludwigsburg	50.50,12.42	
75		*A. quercuscalicis*	Austria	Rosenau	48.00,14.73	
76		*A. quercuscalicis*	Austria	Rosenau	48.00,14.73	
77		*A. quercuscalicis*	Austria	Rosenau	48.00,14.73	
78		*A. quercuscalicis*	Italy	Carru	44.48,7.860	
79		*A. quercuscalicis*	Italy	Carru	44.48,7.860	
80		*A. quercuscalicis*	Italy	Carru	44.48,7.860	
83		*A. quercuscalicis*	Hungary	Godollo	47.60,19.36	
84		*A. quercuscalicis*	Hungary	Godollo	47.60,19.36	
85		*A. quercuscalicis*	Croatia	Istria	45.24,13.87	
86		*A. quercuscalicis*	Croatia	Istria	45.24,13.87	
87		*A. quercuscalicis*	Romania	Sfantu-Gheorghe	45.86,25.78	
87		*A. quercuscalicis*	Slovenia	Ljubljana	46.05,14.51	
88		*A. quercuscalicis*	Slovenia	Ljubljana	46.05,14.51	
89		*A. quercuscalicis*	Romania	Sfantu-Gheorghe	45.86,25.78	
90		*A. quercuscalicis*	Romania	Deva	45.88,22.90	
91		*A. quercuscalicis*	Romania	Sibiu	45.80,24.15	
92		*A. quercuscalicis*	Romania	Sibiu	45.80,24.15	
93		*A. quercuscalicis*	Romania	Sibiu	45.80,24.15	
94		*A. quercustozae*	France	La teste de Buch	44.62,-1.16	
94		*A. quercustozae*	France	Tarbes	43.23,0.080	AY157269
94		*A. quercustozae*	Portugal	Gimonde	41.80,-6.70	
94		*A. quercustozae*	Spain	Barceona	41.40,2.190	AY157269
94		*A. quercustozae*	Spain	Prado del Ray	36.80,-5.55	AY157269
94		*A. quercustozae*	Spain	Prado del Rey	36.79,-5.56	
94		*A. quercustozae*	Spain	Prado del Rey	36.79,-5.56	
94		*A. quercustozae*	Spain	Santiago de Compostela	42.87,-8.55	AY157269
95		*A. quercustozae*	Spain	Madrid	40.42,-3.72	AY157270
96		*A. quercustozae*	Morocco	Azrou	33.45,-5.21	AY157271
97		*A. quercustozae*	France	La teste de Buch	44.62,-1.16	
97		*A. quercustozae*	Morocco	Azrou	33.45,-5.21	AY157272
98		*A. quercustozae*	France	Bordeaux	44.83,-0.57	AY157273
98	1	*A. quercustozae*	Portugal	Gimonde	41.80,-6.70	
99		*A. quercustozae*	France	Perigeux	45.12,0.730	AY157274
100		*A. quercustozae*	Italy	Lame	44.63,9.700	AY157275

cytb	D2	Species	Country	Site	Location	Accession
101		*A. quercustozae*	Italy	Cassina	44.52,10.50	AY157276
102		*A. quercustozae*	Croatia	Istria	45.26,13.94	
102		*A. quercustozae*	Italy	Chiusi	43.03,11.95	AY157277
102		*A. quercustozae*	Italy	Salsomaggiore	44.19,9.620	AY157277
102		*A. quercustozae*	Italy	San Venanzo	42.87,12.27	AY157277
103	1	*A. quercustozae*	Italy	Jelsi	41.53,14.80	AY157278
104		*A. quercustozae*	Italy	Bombiana	44.20,10.95	AY157280
105		*A. quercustozae*	Italy	Greve in Chianti	43.58,11.32	AY157281
106	1	*A. quercustozae*	Italy	Gildone	41.50,14.67	AY157282
106	1	*A. quercustozae*	Italy	Massa Marritima	43.06,10.89	AY157282
107	1	*A. quercustozae*	Greece	Arnissa	40.46,21.56	AY157283
108	1	*A. quercustozae*	Greece	Pisoderi	40.46,21.13	AY157284
109		*A. quercustozae*	Greece	Pisoderi	40.46,21.13	AY157285
110		*A. quercustozae*	Hungary	Lake Balaton	47.10,17.90	AY157286
111		*A. quercustozae*	Greece	Arnissa	40.46,21.56	AY157287
111		*A. quercustozae*	Hungary	Eger	47.88,20.47	AY157287
111	1	*A. quercustozae*	Hungary	Godollo	47.60,19.33	AY157287
111	1	*A. quercustozae*	Hungary	Tatabanya	47.52,18.42	AY157287
112		*A. quercustozae*	Hungary	Veszprem	47.10,17.90	AY157288
114	1	*A. quercustozae*	Turkey	Kulluce	38.20,34.60	AY157289
115	1	*A. quercustozae*	Turkey	Kulluce	38.20,34.60	AY157290
116	1	*A. quercustozae*	Turkey	Yeniyol	41.40,41.63	AY157291
117	1	*A. quercustozae*	Turkey	Refahiye	39.90,38.75	AY157292
117	1	*A. quercustozae*	Turkey	Refahiye	39.90,38.75	
117	1	*A. quercustozae*	Turkey	Refahiye	39.90,38.75	
118		*A. quercustozae*	Portugal	Touça	41.05,-7.23	
119		*A. quercustozae*	Portugal	Touça	41.05,-7.23	
120		*A. quercustozae*	Portugal	Gimonde	41.80,-6.70	
121		*A. quercustozae*	Portugal	Gimonde	41.80,-6.70	
122		*A. quercustozae*	Portugal	Carneiro	41.21,-7.97	
123		*A. quercustozae*	Portugal	Carneiro	41.21,-7.97	
124	1	*A. quercustozae*	Iran	Golestan	33.49,48.35	
125	1	*A. quercustozae*	Iran	Golestan	33.49,48.35	
126	1	*A. quercustozae*	Iran	Golestan	33.49,48.35	
127		*A. quercustozae*	Iran	Golestan	33.49,48.35	
128		*A. quercustozae*	Iran	Golestan	33.49,48.35	
129		*A. quercustozae*	Turkey	Lysandra	36.48,30.05	AY157293
130		*A. quercustozae*	Turkey	Lysandra	36.48,30.05	AY157294
130		*A. quercustozae*	Turkey	Lysandra	36.48,30.05	AY157294
131		*A. quercustozae*	Turkey	Lysandra	36.48,30.05	AY157295
132		*A. quercustozae*	Turkey	Gezende	36.53,33.15	AY157296
133		*A. quercustozae*	Turkey	Gezende	36.53,33.15	AY157297
134		*A. quercustozae*	Turkey	Gezende	36.53,33.15	AY157298
135		*A. quercustozae*	Croatia	Istria	45.26,13.94	
136		*A. quercustozae*	Hungary	Varpalota	47.20,18.14	
137		*A. quercustozae*	Iran	Marivan	35.52,46.17	
138		*A. quercustozae*	Iran	Marivan	35.52,46.17	
139	1	*A. quercustozae*	Iran	Gahvareh	34.35,46.42	
140		*A. quercustozae*	Iran	Gahvareh	34.35,46.42	
141	1	*A. quercustozae*	Iran	Gahvareh	34.35,46.42	
141		*A. quercustozae*	Iran	Gahvareh	34.35,46.42	
141		*A. quercustozae*	Iran	Gahvareh	34.35,46.42	
142	1	*A. quercustozae*	Iran	Gahvareh	34.35,46.42	
143		*A. quercustozae*	Iran	Pyranshahr	36.69,45.14	
144		*A. quercustozae*	Iran	Pyranshahr	36.69,45.14	
145	1	*A. quercustozae*	Spain	Prado del Rey	36.79,-5.56	
145		*A. quercustozae*	Spain	Prado del Rey	36.79,-5.56	
146		*A. dentimitratus*	Hungary	Koszeg	47.39,16.54	
147	1	*A. dentimitratus*	Hungary	Devecser	47.09,17.52	

cytb	D2	Species	Country	Site	Location	Accession
148		*A. dentimitratus*	Hungary	Koszeg	47.39,16.54	
149	1	*A. dentimitratus*	Italy	Siena	43.31,11.33	AF539561
150	1	*A. dentimitratus*	Italy	Foresta Umbra	41.92,15.60	
151	1	*A. dentimitratus*	Italy	Foresta Umbra	41.92,15.60	
152		*A. dentimitratus*	Hungary	Koszeg	47.39,16.54	
152	1	*A. dentimitratus*	Italy	Foresta Umbra	41.92,15.60	
153		*A. dentimitratus*	Italy	Valpiana	43.02,10.84	
154		*A. dentimitratus*	Italy	Valpiana	43.02,10.84	
155		*A. dentimitratus*	Italy	Moio	45.95,9.700	
156		*A. dentimitratus*	Italy	Moio	45.95,9.700	
157	1	*A. dentimitratus*	Spain	Barcelona	41.40,2.190	
157	1	*A. dentimitratus*	Spain	Barcelona	41.40,2.190	
158	1	*A. dentimitratus*	Spain	Barcelona	41.40,2.190	
159	1	*A. dentimitratus*	Spain	Barcelona	41.40,2.190	
160	1	*A. dentimitratus*	Slovenia	Bled	46.37,14.11	
161		*A. dentimitratus*	Turkey	Beybesli	41.03,35.90	
161		*A. dentimitratus*	Turkey	Beybesli	41.03,35.90	
161	1	*A. dentimitratus*	Turkey	Beybesli	41.03,35.90	AF539562
161	1	*A. dentimitratus*	Turkey	Beybesli	41.03,35.90	
161	1	*A. dentimitratus*	Turkey	Beybesli	41.03,35.90	
161	1	*A. dentimitratus*	Turkey	Beybesli	41.03,35.90	
162		*A. dentimitratus*	Turkey	Beybesli	41.03,35.90	
162		*A. dentimitratus*	Turkey	Beybesli	41.03,35.90	
163		*A. dentimitratus*	Turkey	Beybesli	41.03,35.90	
164		*A. dentimitratus*	Turkey	Beybesli	41.03,35.90	
165		*A. pictus*	Spain	Madrid	40.43,-3.70	DQ217997
166		*A. pictus*	Spain	Madrid	40.43,-3.70	DQ217998
167		*A. pictus*	Spain	Madrid	40.43,-3.70	DQ217999
168		*A. coronatus*	Turkey	Küllüce	38.20,34.60	AF539559
169		*A. coronatus*	Italy	Monte Sant'Angelo	42.70,15.92	AF539558
170		*A. coronatus*	Hungary	Sopron	16.59,47.68	AJ228461
171		*A. dentimitratus*	Turkey	Refahiye	39.90,38.75	AF481704
172		*A. mitratus*	Hungary	Godollo	47.60,19.33	AF539567
173		*A. askewi*	Turkey	Cekerek	40.07,35.49	AF539552
174		*A. glutinosus*	Austria	Weinerwald	48.20,16.37	AF539563

Appendix 7

Locations, sample sizes and haplotypes sampled for each population of *Andricus grossulariae*, *Andricus lucidus*, *Andricus megalucidus*, *Biorhiza pallida*, *Cynips quercus* and *Cynips quercusfolii* in Chapter 7 (details for *Andricus coriarius* are presented in Appendix 2). Locations are given in decimal degrees for latitude followed by longitude, and are shown in Figure 7.1. Where multiple copies of a haplotype were sampled from a population, numbers in parentheses after haplotype number indicate the number of individuals sharing that haplotype.

Andricus grossulariae

	Population	Country	Location	Sample size	Haplotypes
1	Istria	Croatia	45.26,13.94	1	34
2	Godollo	Hungary	47.6,19.37	1	13
3	Gahvareh	Iran	34.34,46.41	1	12
4	Piran Shahr	Iran	36.69,45.23	7	2, 3, 34(2), 35, 36, 38
5	Abruzzo	Italy	42.37,13.37	3	4, 19, 30
6	Bombiana	Italy	44.2,10.95	1	28
7	Chiusi	Italy	43.03,11.95	1	20
8	Monte Sant Angelo	Italy	41.92,15.6	2	11, 29
9	Pedemonte	Italy	45.53,11	2	34(2)
10	Poppi	Italy	43.72,11.77	1	4
11	Radicofani	Italy	42.09,11.77	1	27
12	San Venanzo	Italy	42.87,12.27	1	21
13	Azrou	Morocco	33.45,-5.23	6	16, 24, 25, 26, 34(2)
14	Adabia	Spain	41.65,-5.12	2	4(2)
15	Aljucen	Spain	39.03,-6.33	1	5
16	Aglasun	Turkey	37.65,30.53	5	4(2), 34(3)
17	Cekerek	Turkey	40.07,35.49	5	17, 18, 31, 32, 34
18	Egirdir	Turkey	37.87,30.85	3	4(2), 23
19	Gezende	Turkey	36.53,33.15	4	4, 22, 34(2)
20	Kirazoglu	Turkey	39.23,32.44	1	33
21	Ascot	UK	51.41,-0.67	11	4(9), 9, 15

Andricus lucidus*‡/*megalucidus*

	Population	Country	Location	Sample size	Haplotypes
1	Valtice	Czech Republic	48.74,16.76	1	11
2	Foret de Milly	France	48.41,2.7	2	1, 10
3	Nantes	France	47.22,-1.56	2	1(2)
4	Balaton	Hungary	47.1,17.9	1	3
5	Devecser	Hungary	47.09,17.52	1	4
6	Godollo	Hungary	47.6,19.33	1	2
7	Isaszeg	Hungary	47.53,19.4	2	3(2)
8	Jaszbereny	Hungary	47.55,19.99	2	4, 24
9	Kemence	Hungary	40.01,18.88	1	6
10	Sopron	Hungary	47.68,16.59	1	5
11	Szob	Hungary	47.84,18.86	2	1(2)
12	Mátrafüred	Hungary	47.85,19.99	1	30
13	Ghelaie	Iran	33.49,48.35	2	31(2)
14	Golestan	Iran	33.49,48.35	5	19, 20(3), 21
15	Marivan	Iran	35.52,46.17	1	31
16	Piran Shahr	Iran	36.69,45.23	1	23*
17	Pyranshahr	Iran	36.69,45.23	2	G37‡, G39‡
18	Sar-Pol-Zahab	Iran	34.46,45.86	1	31
19	Casina	Italy	44.52,10.5	1	7
20	Castel del Alpi	Italy	44.18,11.28	2	1, 9
21	Chiusi	Italy	43.03,11.95	1	1
22	Cupoli	Italy	42.46,13.83	1	1
23	Gargano	Italy	41.92,15.6	1	8
24	Lame	Italy	44.63,9.7	1	1
25	Pedemonte	Italy	45.53,11	1	1
26	Radicofani	Italy	42.09,11.77	1	1
27	Vernasca	Italy	44.8,9.83	1	1
28	Aandaket St George	Lebanon	34.58,36.3	2	25*, 26*
29	Deir al Zahrani	Lebanon	33.43,35.47	3	G14‡, 28*, 29*
30	Jezzine	Lebanon	33.54,35.59	1	27
31	Aglasun	Turkey	37.65,30.53	1	13*
32	Antalya	Turkey	36.91,30.69	3	1, 14(2)
33	Beysehir	Turkey	36.68,31.73	3	16(3)
34	Cekerek	Turkey	40.07,35.49	1	1
35	Egirdir	Turkey	37.87,30.85	3	1, 12, 15
36	Niksar	Turkey	40.59,36.95	3	11(2), 18
37	Suluova	Turkey	40.84,35.65	1	17
38	Ascot	UK	51.4,-0.67	1	24
39	Kew Gardens	UK	51.47,-0.29	3	22(2), 24
40	Regents Park	UK	51.53,-0.15	7	22, 24(6)

‡ denotes individuals identified as *A. grossulariae*

* denotes individuals identified as *A. megalucidus*

Biorhiza pallida

	Population	Country	Location	Sample size	Halotypes
1	Oberpullendorf	Austria	47.56,16.56	4	1(4)
2	Unterlois	Austria	47.45,16.54	3	1(3)
3	Auxelles	France	47.75,6.77	1	1
4	Clermont-l'Herault	France	43.63,3.43	1	1
5	La Teste de Buch	France	44.63,-1.15	2	1(2)
6	Reims	France	48.72,4.67	5	1(3), 17, 18
7	Rennes	France	48.12,-1.68	3	1, 16(2)
8	St. Jean Pied de Porte	France	43.26,-1.24	4	1(2), 4(2)
9	Ludwigsburg	Germany	9.19,48.9	5	1(5)
10	Bajna	Hungary	47.65,18.6	1	5
11	Godollo	Hungary	47.6,19.33	1	19
12	Matrafured	Hungary	47.85,19.99	1	3
13	Matrahaza	Hungary	47.86,19.98	5	6, 20(4)
14	Szarhalom	Hungary	47.7,16.74	3	1(3)
15	Szeghalom	Hungary	47.23,16.7	1	1
16	Szentendre	Hungary	47.65,19.06	1	1
17	Szentkut	Hungary	47.98,19.78	2	1, 21
18	Golestan	Iran	33.49,48.35	1	26
19	Golestan	Iran	33.49,48.35	1	27
20	Dublin	Ireland	53.34,-6.26	1	6
21	Casina	Italy	44.52,10.5	4	1(2), 7, 25
22	Avila-El Escorial	Spain	40.58,-4.13	1	8
23	Cercedilla	Spain	40.74,-4.05	2	11, 12
24	Guadalix de la Sierra	Spain	40.78,-3.69	2	9, 10
25	Los Molinos	Spain	40.71,-4.07	1	10
26	Prado del Ray	Spain	36.8,-5.55	1	13
27	Puerto de Velate	Spain	43,-1.6	5	1(2), 2, 28, 29
28	Soto del Real	Spain	40.75,-3.78	1	10
29	Yegen	Spain	36.98,-3.13	3	30(3)
30	Zaralejo	Spain	40.55,-4.18	2	10, 11
31	Luin	Switzerland	46.83,9.61	1	1
32	Bolu	Turkey	40.74,31.61	3	27(2), 31
33	Broughton	UK	51.1,-1.6	1	1
34	Cambridge	UK	52.21,0.12	1	1
35	Cannock Chase	UK	52.69,-2.03	1	1
36	Cawood	UK	53.82,-1.13	1	1
37	Hadleigh	UK	51.53,0.6	1	1
38	Knighton Park	UK	52.63,-1.13	2	1(2)
39	Northumberland	UK	55.3,-2.07	1	1
40	Old Sulehey Forest	UK	52.58,-0.42	3	1, 20(2)
41	Oxford	UK	51.75,-1.26	4	1(4)
42	Swaffham	UK	52.63,0.6	1	6

Cynips quercusfolii/quercus†

	Population	Country	Location	Sample size	Haplotypes
1	Wagna	Austria	46.77,15.55	1	32
2	Arlon	Belgium	49.68,5.82	1	2
3	Istria	Croatia	45.26,13.94	1	25
4	Petrovina	Croatia	45.07,16.03	1	4
5	Hodonín	Czech Republic	48.85,17.13	1	4
6	Foret de Signy	France	49.7,4.42	1	4
7	Munich	Germany	48.15,11.45	1	27
8	Chalkidiki	Greece	40.37,23.37	1	6
9	Darány	Hungary	46,17.52	1	7
10	Devecser	Hungary	47.09,17.52	1	21
11	Godollo	Hungary	47.6,19.33	1	4
12	Gyula	Hungary	46.65,21.29	1	23
13	Higher Matrafured	Hungary	47.85,19.99	1	4
14	Pácin	Hungary	48.34,21.83	1	4
15	Sellye	Hungary	45.89,17.83	1	28
16	Sopron	Hungary	47.68,16.59	1	30
17	Kordestan	Iran	35.99,45.9	1	15
18	Lorestan	Iran	33.49,48.35	5	16, 17(2), 18, 19
19	Deir el Zahrani	Lebanon	33.43,35.47	1	20
20	Nijkerk	Netherlands	52.23,5.49	1	4
21	Venray	Netherlands	51.53,5.98	1	31
22	Poland	Poland	50.91,19.31	1	4
23	Celje	Slovenia	46.24,15.27	1	4
24	Ankara	Turkey	39.93,32.86	1	1
25	Bolu	Turkey	41.73,31.61	1	4
26	Hadim	Turkey	38.07,29.6	10	8, 9(2), 10(2), 11, 12, 13, 14, 24
27	Kosedagi	Turkey	41.63,35.37	3	26, 41, 42
28	Bogle Crag	UK	53.02,-3.02	5	3(2), 4(2), 5
29	Claybury	UK	51.61,0.1	1	7
30	Flatts Wood	UK	54.55,-1.92	5	22(5)
31	Loughborough	UK	52.77,-1.21	2	7(2)
32	New Hall	UK	52.21,0.12	2	3, 4
33	Silwood	UK	54.41,-0.67	1	29
34	Wench Ford	UK	51.76,-2.47	5	4(3), 33, 34
35†	Varpalota	Hungary	47.2,18.15	1	46
36†	Arasbaran	Iran	38.6,46.6	7	35(2), 36, 40(2), 47, 48
37†	Khojin-Khalkhal	Iran	37.61,48.53	3	37, 38, 39

† denotes populations identified as *Cynips quercus*

Appendix 8

Cynipid inquiline specimens included in Chapter 8. ISO country codes are used for the localities: AT – Austria, CN – China, DZ – Algeria, ES – Spain, FR – France, GR – Greece, HU – Hungary, IE – Ireland, IR – Iran, IT – Italy, TR- Turkey, UK – United Kingdom, US – United States. The following abbreviations are used for host plants: bra – *Quercus brantii*, cas – *Q. castanaeifolia*, cer – *Q. cerris*, coc – *Q. coccifera*, fag – *Q. faginea*, ile – *Q. ilex*, inf – *Q. infectoria*, pal – *Q. palustris*, pet – *Q. petraea*, pub – *Q. pubescens*, rhu – *Rhus* sp., rob – *Q. robur,* Q. – unidentified *Quercus* species.

species	Mayr Section	code	collecting locality	host plant	host gall inducer	*coxI*	*cytb*	28S D2	28S D3-5
Ceroptres									
C. cerri		S36	Sopron, HU	cer	*Pseudoneuroterus macropterus*	EF486869	EF486975	EF487118	EF487226
C. cerri		S37	Gödöllő, HU	cer	? bud gall	EF486870	EF486976	EF487119	
C. cerri		S166	HU	cer	?		EF486977		
C. clavicornis		S34	Gödöllő, HU	rob	*Andricus conglomeratus*	EF486871	EF486978	EF487120	
C. clavicornis		S35	Gödöllő, HU	pet	*Andricus lignicolus*	EF486872	EF486979	EF487121	EF487227
C. cornigera		S120	Lexington, US	pal	*Challirhytis quercuscornigera*	EF486873	EF486980		
Rhoophilus									
R. loewi		S163	South Africa	rhu	*Scyrotis* sp.	EF486874			
R. loewi		S164	South Africa	rhu	*Scyrotis* sp.	EF486875	EF486981	EF487122	EF487228
R. loewi		S165	South Africa	rhu	*Scyrotis* sp.	EF486876	EF486982	EF487123	EF487229
Saphonecrus									
S. barbotini		S68	Matadapera, ES	coc	*Plagiotrochus britaniae*	EF486877	EF486983	EF487124	EF487230
S. connatus		S50	Mátrafüred, HU	pet	*Andricus testaceipes*	EF486878	EF486984	EF487125	EF487231
S. haimi		S49	Sopron, HU	cer	*Chilaspis nitida*	EF486879	EF486985	EF487126	EF487232
S. irani		S113	Ghelaie, IR	bra	?, leaf gall			EF487127	
S. irani		S114	Ghelaie, IR	bra	?, leaf gall			EF487128	
S. irani		S115	Dar-badam, IR	bra	?, bud gall			EF487129	EF487233
S. irani		S116	Dar-badam, IR	bra	?, bud gall				EF487234
S. lusitanicus		S66	Salamanca, ES	ile	*Plagiotrochus* sp.	EF486880	EF486986	EF487130	EF487235
S. lusitanicus		S67	Salamanca, ES	ile	*Plagiotrochus* sp.	EF486881	EF486987	EF487131	EF487236
S. undulatus		S46	Gödöllő, HU	cer	*Aphelonyx cerricola*	EF486882	EF486988	EF487132	EF487237
S. undulatus		S47	Gödöllő, HU	cer	*Aphelonyx cerricola*	EF486883	EF486989	EF487133	
Synergus									
S. acsi	II.	S112	Ghelaie, IR	bra	*Pseudoneuroterus macropterus*	EF486884	EF486990	EF487134	
S. apicalis, tibialis	II.	S40	Mátrafüred, HU	pet	*Andricus quercusramuli*	EF486885	EF486991	EF487135	
S. apicalis, tibialis	II.	S41	Gödöllő, HU	rob	*Andricus infectorius*	EF486886		EF487136	EF487238
S. apicalis, tibialis	II.	S48	Mátrafüred, HU	pet	*Andricus testaceipes*	EF486887	EF486992	EF487137	EF487239
S. apicalis, tibialis	II.	S51	Mátrafüred, HU	pet	*Andricus quercusramuli*	EF486888	EF486993		EF487240
S. apicalis, tibialis	II.	S52	Mátrafüred, HU	pet	*Andricus testaceipes*	EF486889	EF486994	EF487138	EF487241
S. bechtoldae	II.	S107	Kakareza, IR	inf	*Andricus stonei*		EF486995	EF487139	
S. bechtoldae	II.	S108	Kakareza, IR	inf	*Andricus stonei*		EF486996		
S. chinensis	II.	S90	Mentougou, CH	pet/rob	*Andricus* sp.	EF486890	EF486997	EF487140	
S. clandestinus	I.	S57	Madrid, ES	pyr	?, stunted acorn	EF486891	EF486998	EF487141	EF487242
S. clandestinus	I.	S58	Madrid, ES	pyr	?, stunted acorn	EF486892	EF486999	EF487142	
S. consobrinus	I.	S44	Gödöllő, HU	cer	*Andricus grossulariae*	EF486893		EF487143	
S. consobrinus	I.	S45	Gödöllő, HU	cer	*Andricus grossulariae*	EF486894		EF487144	
S. consobrinus	I.	S55	Gödöllő, HU	cer	*Andricus grossulariae*	EF486895		EF487145	EF487243
S. consobrinus	I.	S56	Gödöllő, HU	cer	*Andricus grossulariae*	EF486896		EF487146	
S. crassicornis	I.	S69	ES	ile	*Plagiotrochus burnayi*	EF486897	EF487000		EF487244

species	Mayr Section	code	collecting locality	host plant	host gall inducer	*coxI*	*cytb*	28S D2	28S D3-5
S. crassicornis	I.	S70	ES	ile	*Plagiotrochus burnayi*	EF486898	EF487001	EF487147	EF487245
S. crassicornis	I.	S132	Matadepera, ES	ile	*Plagiotrochus burnayi*	EF486899			
S. diaphanus	I.	S30	Várpalota, HU	pub	*Andricus conificus*	EF486900	EF487002	EF487148	EF487246
S. diaphanus	I.	S31	Várpalota, HU	pub	*Andricus conificus*	EF486901	EF487003	EF487149	
S. diaphanus	I.	S167	Sopron, HU	pet	*Andricus infectorius*		EF487004		
S. flavipes	I.	S38	Mátrafüred, HU	cer	*Neuroterus lanuginosus*	EF486902	EF487005	EF487150	EF487247
S. flavipes	I.	S39	Szentkút, HU	cer	*Neuroterus saliens*	EF486903	EF487006	EF487151	EF487248
S. flavipes	I.	S130	Sopron, HU	cer	*Pseudoneuroterus macropterus*	EF486904			
S. flavipes	I.	S131	Sopron, HU	cer	*Pseudoneuroterus macropterus*	EF486905	EF487007	EF487152	EF487249
S. gallaepomiformis	II.	S14	Gödöllő, HU	rob	*Andricus hungaricus*	EF486906	EF487008	EF487153	EF487250
S. gallaepomiformis	II.	S15	Gödöllő, HU	rob	*Andricus infectorius*	EF486907	EF487009	EF487154	
S. gallaepomiformis	II.	S168	HU	rob	*Biorhiza pallida*		EF487010		
S. gallaepomiformis	II.	S169	HU	rob	*Andricus kollari*		EF487011		
S. gallaepomiformis	II.	S170	Cercedilla, ES	pyr	*Biorhiza pallida*		EF487012		
S. gallaepomiformis	II.	S171	Szentendre, HU	rob	*Biorhiza pallida*		EF487013		
S. gallaepomiformis, pallicornis	II.	S16	Gödöllő, HU	rob	*Andricus hungaricus*	EF486908	EF487014	EF487155	EF487251
S. gallaepomiformis, pallicornis	II.	S17	Sopron, HU	pub	*Andricus coriarius*	EF486909	EF487015	EF487156	
S. gallaepomiformis, pallipes	II.	S24	Sopron, HU	pub	*Andricus coriarius*	EF486910	EF487016	EF487157	EF487252
S. gallaepomiformis, pallipes	II.	S25	Sopron, HU	pub	*Andricus coriarius*	EF486911	EF487017	EF487158	
S. hayneanus	I.	S8	Szentkút, HU	pet	*Andricus infectorius*	EF486912	EF487018	EF487159	EF487253
S. hayneanus	I.	S9	Várpalota, HU	pub	*Andricus quercustozae*	EF486913	EF487019	EF487160	
S. hayneanus	I.	S77	Gödöllő, HU	pet	*Neuroterus quercusbaccarum*	EF486914	EF487020		
S. hayneanus	I.	S78	Várpalota, HU	rob	*Andricus galeatus*	EF486915	EF487021	EF487161	
S. hayneanus	I.	S79	Gödöllő, HU	rob	*Andricus lignicolus*	EF486916	EF487022	EF487162	
S. hayneanus	I.	S80	Gödöllő, HU	rob	*Andricus kollari*	EF486917	EF487023		
S. hayneanus	I.	S81	Sopron, HU	rob	*Andricus coronatus*	EF486918	EF487024		
S. hayneanus	I.	S172	Karcag, HU	pet/rob	*Andricus lignicolus*		EF487025		
S. hayneanus	I.	S173	Gödöllő, HU	pet/rob	*Andricus lignicolus*		EF487026		
S. hayneanus	I.	S174	Gyöngyös, HU	pet/rob	?		EF487027		
S. hayneanus	I.	S175	Veszprém, HU	pet/rob	*Andricus quercustozae*		EF487028		
S. hayneanus, reinhardi	I.	S6	Várpalota, HU	pub	*Andricus quercustozae*	EF486919	EF487029	EF487163	EF487254
S. hayneanus, reinhardi	I.	S7	Szentkút, HU	pub	*Andricus caputmedusae*	EF486920	EF487030	EF487164	
S. hayneanus, reinhardi	I.	S73	Szentkút, HU	pub	*Andricus caputmedusae*	EF486921			

species	Mayr Section	code	collecting locality	host plant	host gall inducer	*coxI*	*cytb*	28S D2	28S D3-5
S. hayneanus, reinhardi	I.	S74	Gödöllő, HU	rob	*Andricus grossulariae*	EF486922			
S. hayneanus, reinhardi	I.	S75	Várpalota, HU	pub	*Andricus quercustozae*	EF486923	EF487031		
S. hayneanus reinhardi	I.	S76	Mátrafüred, HU	pub	*Andricus caputmedusae*	EF486924	EF487032		
S. incrassatus	II.	S59	Madrid, ES	pyr	*Andricus quercusradicis*	EF486925	EF487033	EF487165	EF487255
S. incrassatus	II.	S176	HU	pet/rob	*Andricus glutinosus*		EF487034		
S. japonicus	I.	S92	JP	Q.	*Andricus* sp.	EF486926	EF487035	EF487166	
S. japonicus	I.	S96	JP	Q.	*Andricus* sp.	EF486927	EF487036	EF487167	
S. japonicus	I.	S97	JP	Q.	*Andricus* sp.			EF487168	
S. mikoi	I.	S105	Ghelaie, IR	inf	*Andricus chodjaii*	EF486928	EF487037	EF487169	
S. mikoi	I.	S106	Ghelaie, IR	inf	*Andricus chodjaii*	EF486929	EF487038		
S. pallicornis	II.	S20	Szentkút, HU	pet	*Andricus infectorius*	EF486930	EF487039	EF487170	
S. pallicornis	II.	S21	Gödöllő, HU	rob	*Andricus caputmedusae*	EF486931		EF487171	EF487256
S. pallicornis	II.	S178	Balaton, HU	pub	*Andricus quercustozae*		EF487040		
S. pallicornis	II.	S179	HU	pet/rob	*Andricus coriarius*		EF487041		
S. pallicornis, gallaepomiformis	II.	S18	Gödöllő, HU	rob	*Andricus hungaricus*	EF486932	EF487042	EF487172	
S. pallicornis, gallaepomiformis	II.	S19	Mátrafüred, HU	pet	*Andricus infectorius*	EF486933	EF487043	EF487173	EF487257
S. pallicornis, pallipes	II.	S22	Gödöllő, HU	rob	*Andricus infectorius*	EF486934	EF487044	EF487174	
S. pallicornis, pallipes	II.	S23	Sopron, HU	pub	*Andricus coriarius*	EF486935		EF487175	EF487258
S. pallicornis, pallipes	II.	S88	Gödöllő, HU	rob	*Andricus lignicolus*	EF486936			
S. pallicornis, pallipes	II.	S89	Sopron, HU	pub	*Andricus coriarius*	EF486937			
S. pallidipennis	I.	S12	Gödöllő, HU	rob	*Andricus hungaricus*	EF486938	EF487045	EF487176	EF487259
S. pallidipennis	I.	S13	Várpalota, HU	pub	*Andricus conificus*	EF486939	EF487046	EF487177	EF487260
S. pallidipennis	I.	S181	HU	pet/rob	*Andricus kollari*		EF487047		
S. pallipes	II.	S26	Sopron, HU	pub	*Andricus coriarius*	EF486940		EF487178	EF487261
S. pallipes	II.	S27	Gödöllő, HU	cer	*Andricus hungaricus*	EF486941	EF487048	EF487179	
S. pallipes	II.	S62	Madrid, ES	pet	*Andricus quadrilineatus*	EF486942	EF487049	EF487180	
S. pallipes	II.	S63	Madrid, ES	pet	*Andricus quadrilineatus*	EF486943	EF487050	EF487181	EF487262
S. pallipes	II.	S82	Sopron, HU	pub	*Andricus coriarius*		EF487051		
S. pallipes	II.	S83	Szentkút, HU	pub	*Neuroterus anthracina*	EF486944		EF487182	
S. pallipes	II.	S86	Sopron, HU	pub	*Andricus coriarius*	EF486945			
S. pallipes	II.	S87	Mátrafüred, HU	pet	*Neuroterus albipes*	EF486946	EF487053		
S. pallipes	II.	S177	Sopron, HU	pet/rob	*Andricus kollari*		EF487054		
S. palmirae	II.	S128	Caspean sea, IR	cas	*Neuroterus* sp.	EF486947			
S. physocerus	II.	S28	Várpalota, HU	rob	*Trigonaspis synaspis*	EF486948	EF487055	EF487183	EF487263
S. physocerus	II.	S29	Várpalota, HU	rob	*Trigonaspis synaspis*	EF486949	EF487056	EF487184	

species	Mayr Section	code	collecting locality	host plant	host gall inducer	*coxI*	*cytb*	28S D2	28S D3-5
S. physocerus	II.	S60	Madrid, ES	fag	*Trigonaspis synaspis*	EF486950	EF487057	EF487185	EF487264
S. physocerus	II.	S61	Madrid, ES	fag	*Trigonaspis synaspis*	EF486951	EF487058	EF487186	
S. plagiotrochi	II.	S64	Madrid, ES	ile	*Plagiotrochus australis*	EF486952		EF487187	EF487265
S. plagiotrochi	II.	S65	Madrid, ES	ile	*Plagiotrochus australis*	EF486953	EF487059	EF487188	
S. reinhardi	I.	S182	Gödöllő, HU	pet/rob	*Andricus lignicolus*		EF487060		
S. reinhardi	I.	S183	Balaton, HU	pet/rob	*Andricus caputmedusae*		EF487052		
S. reinhardi	I.	S184	Gödöllő, HU	pet/rob	*Andricus caputmedusae*		EF487061		
S. reinhardi	I.	S185	Eger, HU	pet/rob	*Andricus caputmedusae*		EF487062		
S. reinhardi	I.	S186	Egirdir, TR	inf	*Andricus caputmedusae*		EF487063		
S. reinhardi	I.	S187	Vienna, AT	pet/rob	*Andricus caputmedusae*		EF487064		
S. reinhardi	I.	S188	Balaton, HU	pet/rob	*Andricus coronatus*		EF487065		
S. reinhardi	I.	S189	Szentendre, HU	pet/rob	*Andricus glutinosus*		EF487066		
S. reinhardi	I.	S190	Veszprém, HU	pet/rob	*Andricus quercustozae*		EF487067		
S. reinhardi	I.	S191	Veszprém, HU	pet/rob	*Andricus quercustozae*		EF487068		
S. reinhardi	I.	S192	Veszprém, HU	pet/rob	*Andricus quercustozae*		EF487069		
S. reinhardi	I.	S193	Veszprém, HU	pet/rob	*Andricus quercustozae*		EF487070		
S. reinhardi	I.	S194	Veszprém, HU	pet/rob	*Andricus quercustozae*		EF487071		
S. thaumacerus	II.	S42	Gödöllő, HU	cer	*Andricus grossulariae*	EF486954		EF487189	
S. thaumacerus	II.	S43	Sopron, HU	cer	*Chilaspis nitida*	EF486955		EF487190	EF487266
S. thaumacerus	II.	S53	Gödöllő, HU	cer	*Neuroterus saliens*	EF486956	EF487072	EF487191	
S. thaumacerus	II.	S54	Gödöllő, HU	cer	*Neuroterus saliens*	EF486957	EF487073	EF487192	EF487267
S. umbraculus	I.	S1	Sopron, HU	pet	*Andricus infectorius*	EF486958	EF487074	EF487193	EF487268
S. umbraculus	I.	S2	Mátrafüred, HU	pub	*Andricus lucidus*	EF486959	EF487075	EF487194	
S. umbraculus	I.	S3	Sopron, HU	rob	*Andricus quercuscalicis*		EF487076		
S. umbraculus	I.	S5	Gödöllő, HU	rob	*Andricus hungaricus*	EF486960	EF487077	EF487195	
S. umbraculus	I.	S100	IR	inf	?	EF486961	EF487078	EF487196	
S. umbraculus	I.	S101	IR	inf	?	EF486962	EF487079		
S. umbraculus	I.	S102	Ghelaie, IR	inf	*Andricus caputmedusae*	EF486963	EF487080	EF487197	
S. umbraculus	I.	S103	Ghelaie, IR	inf	*Andricus caputmedusae*	EF486964	EF487081		
S. umbraculus	I.	S104	Ghelaie, IR	inf	*Andricus quercustozae*			EF487198	
S. umbraculus	I.	S139	Debrecen, HU	rob	*Andricus foecundatrix*		EF487082	EF487199	
S. umbraculus	I.	S140	Debrecen, HU	rob	*Andricus foecundatrix*		EF487083	EF487200	
S. umbraculus	I.	S141	Debrecen, HU	rob	*Andricus foecundatrix*		EF487084	EF487201	
S. umbraculus	I.	S142	Gödöllő, HU	rob	*Andricus polycerus*		EF487085		
S. umbraculus	I.	S143	Gödöllő, HU	rob	*Andricus kollari*		EF487086	EF487202	
S. umbraculus	I.	S144	Gödöllő, HU	rob	*Andricus kollari*		EF487087		
S. umbraculus	I.	S145	Gödöllő, HU	rob	*Andricus polycerus*		EF487088	EF487203	
S. umbraculus	I.	S146	Várpalota, HU	pub	*Andricus polyecrus*		EF487089	EF487204	
S. umbraculus	I.	S147	Borsod, HU	pet/rob	*Andricus lignicolus*		EF487090		
S. umbraculus	I.	S148	Várpalota, HU	pub	*Andricus polycerus*		EF487091		

species	Mayr Section	code	collecting locality	host plant	host gall inducer	*coxI*	*cytb*	28S D2	28S D3-5
S. umbraculus	I.	S149	Várpalota, HU	pub	*Andricus kollari*		EF487092	EF487205	
S. umbraculus	I.	S150	Debrecen, HU	pet/rob	*Andricus infectorius*		EF487093		
S. umbraculus	I.	S151	Várpalota, HU	pet/rob	*Andricus kollari*		EF487094		
S. umbraculus	I.	S152	Debrecen, HU	pet/rob	*Andricus infectorius*			EF487206	
S. umbraculus	I.	S154	Gödöllő, HU	pet/rob	*Andricus coriarius*			EF487207	
S. umbraculus	I.	S155	Gödöllő, HU	pet/rob	*Andricus infectorius*			EF487208	
S. umbraculus	I.	S156	Debrecen, HU	pet/rob	*Andricus lignicolus*			EF487209	
S. umbraculus	I.	S157	Borsod, HU	pet/rob	*Andricus lignicolus*			EF487210	
S. umbraculus	I.	S158	Gödöllő, HU	pet/rob	*Andricus kollari*			EF487211	
S. umbraculus	I.	S159	Gödöllő, HU	pet/rob	*Andricus kollari*			EF487212	
S. umbraculus	I.	S160	Gödöllő, HU	pet/rob	*Andricus kollari*			EF487213	
S. umbraculus	I.	S161	Gödöllő, HU	pet/rob	*Andricus kollari*			EF487214	
S. umbraculus	I.	S162	Várpalota, HU	pub	*Andricus polycerus*			EF487215	
S. umbraculus	I.	S196	Cercedilla, ES	Q.	*Biorhiza pallida*		EF487095		
S. umbraculus	I.	S197	Guerande, FR	Q.	*Andricus kollari*		EF487096		
S. umbraculus	I.	S198	Cercedilla, ES	Q.	*Biorhiza pallida*		EF487097		
S. umbraculus	I.	S199	Cercedilla, ES	Q.	*Biorhiza pallida*		EF487098		
S. umbraculus	I.	S200	London, UK	pet/rob	*Biorhiza pallida*		EF487099		
S. umbraculus	I.	S201	HU	pub	*Andricus polycerus*		EF487100		
S. umbraculus	I.	S202	HU	cer	*Neuroterus saliens*		EF487101		
S. umbraculus, hayneanus	I.	S10	Szentkút, HU	pub	*Andricus coriarius*	EF486965	EF487102	EF487216	
S. umbraculus, hayneanus	I.	S11	Gödöllő, HU	rob	*Andricus lucidus*	EF486966	EF487103	EF487217	EF487269
S. variabilis	II.	S98	Nogian, IR	bra	*Neuroterus lanuginosus*			EF487218	
S. variabilis	II.	S99	Nogian, IR	bra	*Neuroterus lanuginosus*	EF486967	EF487104	EF487219	
S. variabilis	II.	S195	Veszprém, HU	cer	*Aphelonyx cerricola*		EF487105		
S. xiaolongmeni	II.	S94	Mentougou, CH	rob	*Andricus* sp.	EF486968	EF487106	EF487220	
S. sp.		S109	Ghelaie, IR	bra	*Aphelonyx persica*	EF486969	EF487107	EF487221	
S. sp.		S138	Sopron, HU	cer	*Chilaspis nitida*	EF486970		EF487222	EF487270
S. sp.		S203	HU	rob	*Andricus legitimus*		EF487108		
S. sp.		S204	IE	rob	*Andricus foecundatrix*		EF487109		
Synophrus									
S. politus		S32	Sopron, HU	cer		EF486971	EF487110	EF487223	EF487271
S. pilulae		S33	Sopron, HU	cer		EF486972	EF487111	EF487224	
S. sp. nova		S134	DZ	sub		EF486973			EF487272
S. sp. nova		S135	ES	Q.		EF486974		EF487225	
S. politus		S205	Püspökladány, HU	cer			EF487112		
S. politus		S206	Sopron, HU	cer			EF487113		
S. politus		S207	Lame, IT	cer			EF487114		

species	Mayr Section	code	collecting locality	host plant	host gall inducer	*coxI*	*cytb*	28S D2	28S D3-5
S. politus		S208	Madenli, TR	Q.			EF487115		
S. politus		S209	Madenli, TR	Q.			EF487116		
S. politus		S210	Edessa, GR	Q.			EF487117		